Enhancing the Appraisal of Acute Mental Health Crisis: The Crisis Risk and Adaptive Functioning Tool (CRAFT)

NICOLE YVETTE STOKOE

Submitted in partial fulfilment of the requirements of the University of Hertfordshire for the degree of Doctor of Philosophy

September 2012

Acknowledgements

I would like to express my warm thanks to Professor Gary Kupshik and Professor Ben Fletcher for all their encouragement, guidance, help and support over the course of this research. In particular I would like to thank Professor Gary Kupshik for his continued faith in me throughout, for helping me to overcome what felt to be insurmountable obstacles and for being the clinical perspective of the project. I would also like to thank Professor Ben Fletcher for his great ideas, ability to think outside the box and helpful challenges that expanded the scope and reach of the research.

I would also like to thank the Bedford and Luton Crisis Resolution and Home Treatment teams and South Essex Partnership University NHS Foundation Trust for their participation and contributions without which this project would not have been able to move forwards.

For their continued statistical genius and insights I would like to thank Sam Norton and Mike Horton who were always quick to respond to my statistical cries for help. Their continued support and willingness to share their extensive knowledge provided much needed reassurance and guidance at critical points in the research journey.

Finally, I would like to thank my family and friends; especially my mum Eve, who took the time and patience to painstakingly read through the entirety of this thesis and as a consequence has been subjected to Rasch analysis, for which I can only apologise.

Table of Contents

List of Figures

List of Tables

Table of Appendices

Abstract

There is a distinct lack of research into the concept of acute mental health crisis. Without investigating the concept of crisis itself, it is not possible to appreciate the attributes of crisis so that it can be measured. This has hampered the development of good psychometric tools for crisis.

The aim of this research was to develop the first standardised, valid and reliable measure for the assessment of people presenting to Crisis Resolution and Home Treatment (CRHT) teams. This research utilised qualitative and quantitative research techniques to develop a crisis measure starting with a comprehensive investigation into the concept of acute mental health crisis to identify an item pool and clinically credible item rating scale. A prototype crisis measure was developed and piloted in two NHS CRHTs and data collected. This data was analysed to identify the key areas of crisis assessment (the subscales), a flexible rating scale and scoring system creating a measure named the Crisis Risk and Adaptive Functioning Tool (CRAFT). The CRAFT provides patient crisis profiles highlighting areas of strength, resilience, weakness and vulnerability.

A thorough investigation of crisis was completed with CRHT staff and patients through interviews and focus groups. An initial 143 item pool and clinically credible item rating scale were identified and developed into a prototype pilot crisis measure. This measure utilised a flexible rating approach encapsulating both risk and protective factors believed to reflect clinical practice.

The measure was piloted and the data analysed to assess the structure of the crisis measure's item pool using the statistical techniques of Principal Component Analysis (PCA) and Rasch analysis. These analyses resulted in a 66 item measure with 8 unidimensional subscales including; 1) Crisis Recovery Indicators, 2) Adaptive Decision Making, 3) Risk of Harm to Self, 4) Mediating Factors, 5) Daily Structure, 6) Risk of Harm to Others, 7) Feelings and Affect, and 8) Basic Needs. The total variance explained by these 8 subscales was 67.6% with excellent internal reliability as indicated by a Cronbach's alpha coefficient of 0.98 ($p<0.001$) and temporal reliability indicated by Spearman's correlation of 0.971 ($p<0.001$, one tailed). This suggests that this measure has a strong internal structure and provides stable outcomes over time at both the subscale and global overall measurement levels.

Receiver Operator Characteristic curve analysis supported the identification of cut-offs to indicate low, moderate and high levels of crisis and were shown to have good levels of sensitivity and specificity for the crude discrimination between individuals who require CRHT treatment and individuals who do not require CRHT treatment (sensitivity 0.89 and specificity 0.73) and for accurately discriminating between the basic treatment levels of low, moderate and high (sensitivity 0.80; specificity 0.69).

One of the great advantages of utilising the Rasch model is that it supports the identification of key characteristics from an item pool. Application of the subscales and the overall measure to the Rasch model identified items that were most representative of underlying constructs and risk, highlighting items of essential essence for assessing crisis in the context of community treatment. These items may act as useful clinical and risk indicators for community assessment.

After considering the evidence from the PCA and Rasch analysis for the underpinning construct, the measure was named the Crisis Risk and Adaptive Functioning Tool (CRAFT) to encapsulate both the risk and adaptive functioning (coping and management) aspects measured by the tool.

There are a number of clinical implications resulting from the development of CRAFT for the assessment of crisis. This research clearly identifies 8 key areas for crisis assessment and the specific items that describe them. This promises to be a powerful clinical tool as it clarifies the main areas of concern and importance for crisis assessment and provides mental health professionals with a means of assessing and monitoring patients experiencing crisis. In addition to the clinical benefits offered by the CRAFT, it provides an approach to assessing and monitoring crisis to support further research in the area of acute mental health crisis.

This research offers significant steps towards the development of a quality measure for crisis assessment. However, it is acknowledged that the process of measurement development is never complete. It simply evolves over time with the aim of coming closer to the valued direction.

Chapter 1

Literature Review

1.0 Understanding the Nature of Acute Mental Health Crisis

1.0.1 Introduction

The aim of this research is to develop a standardised, valid and reliable measure to assist mental health professionals assessing individuals experiencing acute mental health crises.

This chapter is a review of the relevant literature. The chapter sets the foundation on which this research is based and also contributes to developing a comprehensive understanding of mental health crisis. It is important to fully understand and conceptualise the nature of crisis in order to identify its antecedents and make appropriate clinical decisions. The nature and make up of crises are extremely complex and personal to each person, affected by a number of external and internal variables. Without fully realising what the crisis variables are, it will be impossible to measure. Identification of the core characteristics of the concept of crisis is a crucial first step.

One of the challenges for measuring crisis is that it cannot be directly observed and is therefore described as a latent variable or construct. A latent variable or construct is a thing or aspect of a person that cannot be directly observed (Nunnally & Bernstein, 1994) but must be inferred from the observation of other indicators that are believed to directly relate to the underlying construct of interest. This chapter looks at the literature that describes acute mental health crisis as a means toward developing a working understanding or description of the underlying latent construct or constructs of crisis that this research aims to measure.

Following a comprehensive exploration of the crisis definition there will be a description of associated theories that contribute to the current understanding of crisis, models of crisis intervention, current treatment approaches and the role of assessment and measurement. This literature review will provide information for understanding crisis and will also provide evidence for the rationale behind this research that calls for the development of a crisis measure.

1.0.2 Understanding and Defining Acute Mental Health Crisis

This section aims to describe the history and roots of crisis theory, the theoretical underpinnings that have contributed to our understanding of crisis and to summarise this into a current working definition of crisis from which this research will grow.

1.0.2.1 Historical Overview of Crisis

'Acute mental health crisis' is the full term for describing what will be simply referred to as crisis from hereon after. One of the first attempts to define the state of crisis was made by Thomas (1909). He defined it as 'a threat, a challenge, a strain on the attention, a call to new action, which may have the germ of a new organisation'. This definition is basic in terms of its ability to fully explain the triggers, experience and consequences of the crisis state but the key concepts stated by Thomas are still present in the current crisis theoretical understanding and literature. As outlined in section 1.0.4.3, it is expected that as an individual develops from infancy to adulthood, they will encounter a number of developmental or expected life crises (Erikson, 1968) which would not be considered a mental health crisis but may act as a trigger to this experience. By comparison to the developmental crises described by Erikson, mental health crises are unexpected in nature.

1.0.2.2 The Founders of Crisis Theory - Lindemann (1944) and Caplan (1961)

The foundations of current crisis theory lie in the work of Lindemann (1944; 1956). Lindemann and his colleagues were some of the first mental health professionals to recognise that everyone is vulnerable to crisis regardless of background or history. The concept of dealing with non-developmental, unexpected or situational life crisis first described by Lindemann (1944) in relation to acute grief reactions was used later by Caplan (1961) to the wider spectrum of crises. The types of crisis explained through the work of Lindemann and Caplan are unexpected, with sudden onset, and are generally out of the control of the individual experiencing the crisis (Kanel, 1998; Slaideu, 1990).

What was particularly ground-breaking at the time was the suggestion that the experience of crisis was not indicative of significant pathology but reflected the individual striving to negotiate an unexpected obstacle to a life goal or goals. The emphasis here is that enduring mental health pathology would not *necessarily* result as a direct consequence of the experience of crisis. However, it is important to acknowledge that individuals with pre-existing mental health difficulties may be more vulnerable to experiencing crisis and without successful crisis resolution an individual may become more vulnerable to developing enduring mental health difficulties in the future. Lindemann simply emphasised that these were not necessary attributes of the crisis presentation and therefore crisis could be equally observed in an individual experiencing a significant relationship breakdown or financial difficulties as it could be observed in an individual who had a long term history of depression and is experiencing a breakdown in ability to manage this mental health difficulty.

It was Lindemann's (1944) clinical observations that led to his theory around acute grief as a specific example of situational crises. Lindemann observed that similar symptoms and pathways to recovery presented in all of the observed cases where acute grief and trauma were experienced. He noted that the crisis period of the acute grief reaction was approximately 4-6 weeks and suggested that the treatment of crisis should be focused, short term and time-limited with the goal being to return the individual to their home as soon as possible rather than retaining them in hospital. At the time, these suggestions would have been considered an innovative and bold approach. To put this in context, Lindemann was endorsing a community treatment approach for individuals experiencing crisis at a time when individuals could be admitted to psychiatric institutions for simply causing difficulties in their family relationships at home, for being elderly, or for being pregnant out of wedlock (Grob, 1992, Morton, 1937). Lindemann suggested that community treatment and management of crisis should be supported by either professionals or a combination of family and/or community members and professionals: "The help of a social worker or a minister, or if these are not available, a member of the family, to urge the patient to continue coming to see the psychiatrist may be indispensable……….Social workers and ministers will have to be on the look-out for the more ominous pictures, referring these to the psychiatrist while assisting the more normal reactions themselves." (p.147). This statement acknowledges the importance of social networks for support and the importance of protective factors in the community for aiding successful recovery from crisis.

Caplan (1961) translated the work of Lindemann and expanded the concept to the wider spectrum of crisis. One of the most influential contributors to crisis theory, Caplan (1961; p18) defined crisis as "…….provoked when a person faces an obstacle to important life goals that is, for a time, insurmountable through the utilisation of customary methods of problem-solving. A period of disorganization ensues, a period of upset, during which many different abortive attempts at solution are made. Eventually some kind of adaptation is achieved, which may or may not be in the best interests of that person and his fellows." This definition pointed toward coping theory (Lazarus & Folkman, 1984) with the idea of a breakdown in ability to problem solve (to cope) and as a consequence to find a solution for obstacles to life goals. It also introduced the idea that the resolution of crisis may or may not be a positive or helpful one.

Caplan related crisis to the concept of homeostasis (detailed in section 1.0.4.1), in terms of an individual's desire to maintain a state of sameness, stasis or a state of stability and equilibrium. Similarly to physiological homeostasis, when the individual is moved out of their stable state, they will strive to regain psychological homeostasis through the implementation of previously successful coping strategies. Where attempts at stasis fail, crisis results.

The work of Lindemann and Caplan provided a useful foundation to crisis theory which has subsequently been expanded and applied in clinical practice.

1.0.2.3 Developing Crisis Theory Towards an Integrated Understanding – Hobbs to Roberts

The Yerke's-Dodson's law (1908) outlines the relationship between arousal and performance indicating that too little or too much arousal results in hindered performance. Too little arousal and an individual is essentially 'bored' which prevents good performance, therefore as human beings it is preferable to experience low levels of pressure or arousal in order to improve performance. Performance increases with an increase in arousal until the maximal level is reached where an individual is experiencing the optimum amount of arousal to produce optimum performance levels. However, if arousal increases far beyond this point, the individual is no longer able to manage and performance starts to deteriorate until the point the person becomes completely overwhelmed and essentially burns out. Hobbs (1984) made a useful comparison between the Yerkes-Dodson's arousal curve and Caplan's (1964b) four phases of crisis development by developing a Crisis Curve (Figure 1.0). Caplan described crisis as occurring over four phases with an initial rise in tension resulting from an obstruction to a life goal, increased disruption due to an inability to overcome this obstacle, continued increasing tension as emergency problem-solving methods failed, which may finally result in a breakdown or crisis state if the individual is unable to exit the process. Hobb's suggested that at the third phase of Caplan's model the individual may present with 'maximal resourcefulness' as they continue to utilise previously known coping strategies as well as attempting new coping approaches, forced to try coping approaches outside of their normal coping repertoire. At each phase, where the individual is successful in finding approaches to cope with the obstacles to their life goals, arousal levels may start to decrease, returning back to normal levels. However, where an individual is unsuccessful they will progress eventually through the next phases, bringing them closer to the experience of crisis, challenging their coping ability and resilience. This model presents the idea that when an individual's level of arousal is too high, their ability to cope or to be resourceful is significantly challenged and may result in the overall breakdown in coping which will be experienced as crisis (Figure 1.0). It is important to note that no time line or specified direction of phase transition has been applied to this model. This represents the individual nature of crisis whereby individuals can move between the phases of crisis in either direction and over varying timelines.

Figure 1.0 Hobbs (1984) Crisis Curve

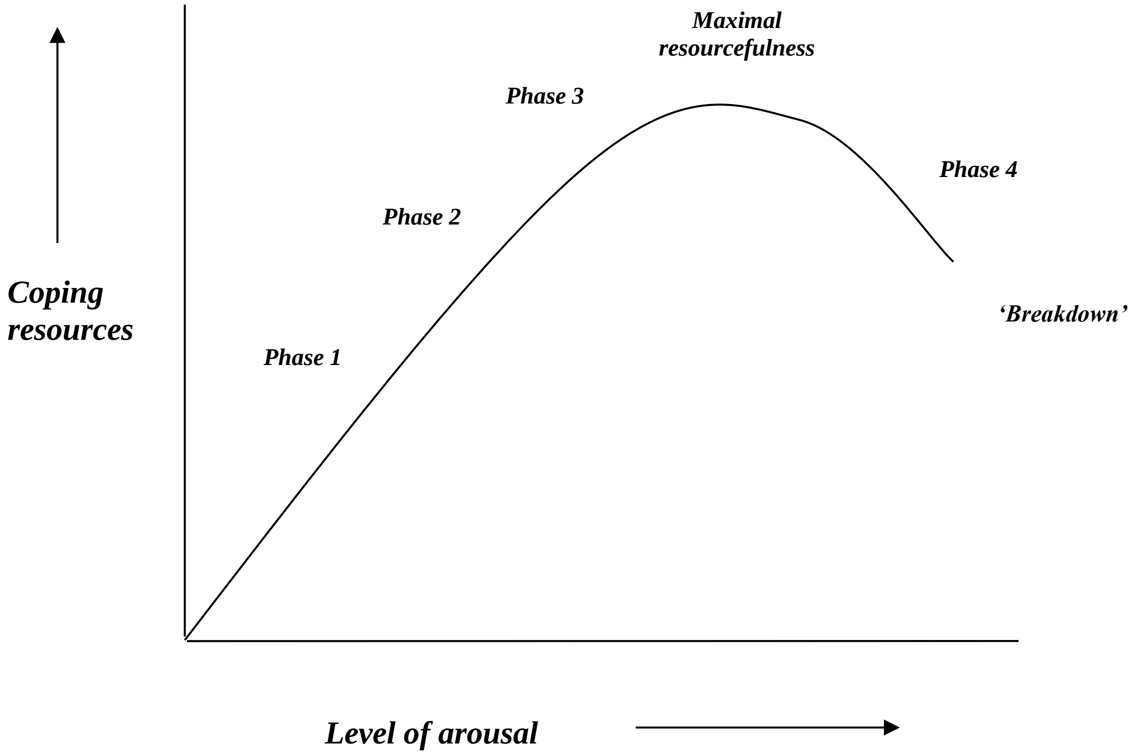


Figure 1.0: The Hobbs crisis curve (1984, p.28) depicting the relationship between increasing levels of arousal and the utilisation of coping resources. The 'maximal resourcefulness' point on the curve indicates the point where arousal levels challenge the individual to utilise all of their current coping strategies, possibly stretching their coping repertoire to try new coping approaches.

Hobbs (1984, p.29) described the characteristics of crisis as summarised below:

1. Crises are self-limiting. Some kind of resolution to crisis will be found within 4-6 weeks. This can be a helpful or unhelpful, an adaptive or maladaptive resolution of the crisis.
2. Dependency needs are expressed at the early stages of crisis although these may not always be communicated directly.
3. Individual crisis may reflect crisis on a larger scale for example within a family or social system.
4. Crisis is not, in itself, a pathological state.
5. Crisis may present the opportunity for resolution of old conflicts, derived from the maladaptive solution of earlier crises.

Hobbs' (1984) crisis definition is particularly helpful for outlining the role of the wider society and community, integrating previous theory by Erikson (1968) which suggests that previous maladaptive resolution of crisis likely to impact on the person's ability to successfully resolve future crises. In addition, Hobbs (1984) outlined the relationship between protective factors, vulnerability factors and crisis resolution as depicted here in Figure 1.1:

Figure 1.1 Hobbs (1984) Crisis Model

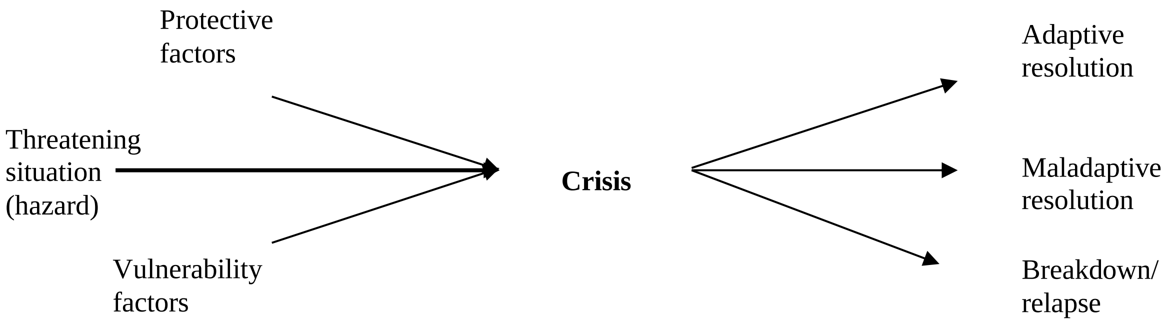


Figure 1.1: Hobbs (1984, p.27) crisis model showing the relationship between protective and vulnerability factors and the possible outcomes for the resolution of crisis as a result of the precipitating factors.

Kanel (1998) saw crisis as the sum of three parts; "1) a **precipitating event** occurs, 2) the perception of this event leads to **subjective distress** 3) usual **coping methods fail**, leaving the person experiencing the event to function psychologically, emotionally, or behaviourally at a lower level than before the precipitating event" (p.1). Therefore his definition of crisis has been termed the trilogy definition with the interesting aspect that it outlines the importance of individual perception. It is the interpretation of the stress factor or the pressure rather than the stress factor itself that precipitates crisis. Theory relating to cognitive appraisal, perception and interpretation is summarised in section 1.0.4.4 and clearly links in with coping theory. Kanel also clearly states here that where coping methods fail, the individual will consequently function at a lower level than their pre-crisis state in the three areas of psychological, emotional and behavioural functioning which links in with previous work by Maslow (1943) and the concept of the hierarchy of needs (section 1.0.4.2).

It became fashionable to relate the experience of crisis to the Chinese symbol for crisis. This symbol is made up of two words, *danger* and the other representing *opportunity*. For example, Hoff (2001) defined a crisis as ‘a serious occasion or turning point presenting both danger and opportunity.’(p.4). Crisis previously held purely negative connotations. However, this definition reflects the shift in thinking that occurred which resulted in a more flexible understanding of this presentation. This new conceptualisation recognises crisis as a juncture that can take the person in a number of directions, both positive and negative. How crisis is resolved is key to the crisis outcome in terms of future functioning and vulnerability to further crises or the development of enduring mental health difficulties. As the window of crisis is brief (approximately 4-6 weeks), intervention must be timely and focused to give the best possible chance of a positive recovery outcome.

James & Gilliland (2001) focused down the broad concept of crisis into four specific crisis areas, influenced by the theories of Erikson (1968, section 1.0.4.3), Maslow (1943, section 1.0.4.2) and Lazurus & Folkman’s theory of coping (1984, section 1.0.4.4). Firstly they described the concept of developmental crises, crises that are the expected normal crises that need to be negotiated as a part of the life journey. Secondly they described situational crises, which are uncommon and extraordinary events that cannot be predicted by the individual. For example; a car crash, an unexpected death of a loved one, rape, the loss of a job etc. “The key to differentiating a situational crisis from other crises is that a situational crisis is random, sudden, shocking, intense and often catastrophic.” (James & Gilliland, 2001, p.5). The third type of crisis explored was existential crises defined as inner conflicts and anxieties that accompany important human issues of purpose, responsibility, independence, freedom, and commitment. The final types of crisis explored by James and Gilliland were environmental crises. These are where either a natural or human made event occurs that is catastrophic to a level that all persons involved in that environment are affected and have to deal with the consequences. For example, the impact of earthquakes, famine or tsunami

In general, the types of crisis that imply acute mental health crisis are often situational or existential crisis. However crisis may also be triggered by developmental and environmental factors. These acute episodes are supported in a number of countries by crisis teams made up of multidisciplinary mental health professionals whose role is to guide individuals toward healthy crisis resolution. In the United Kingdom, these teams are called CRHT teams and are the target users for the crisis measure to be developed through this research. CRHTs are described in more detail in sections 1.0.6 and 1.0.7.

Roberts (2005, p.94) integrated previous theory to develop one of the most widely used definitions of acute mental health crisis in current use today:

> "An acute disruption of psychological homeostasis in which one's usual coping mechanisms fail and there exists evidence of distress and functional impairment. The subjective reaction to a stressful life experience that compromises the individual's stability and ability to cope or function. The main cause of a crisis is an intensely stressful, traumatic, or hazardous event, but two other conditions are also necessary: (1) the individual's perception of the event as the cause of considerable upset and/or disruption; and (2) the individual's inability to resolve the disruption by previously used coping mechanisms. Crisis also refers to 'an upset in the steady state'. It often has five components: a hazardous or traumatic event, a vulnerable state, a precipitating factor, an active crisis state, and the resolution of the crisis."

This definition encapsulates the current key concepts for understanding acute mental health crisis developed over the last century. This includes ideas around coping, triggers, perception, obstacles to life goals, stress or pressure, an individual's drive to maintain equilibrium and crisis resolution.

1.0.2.4 Developing Crisis Theory in the Modern Context

Tobitt & Kamboj (2011) identified the limitations of the crisis ‘definition’ offered by the Department of Health’s (2001) policy. They noted that it was unable to characterise the essential features of the crisis presentation and therefore CRHTs lacked the guidance of a strong theoretical foundation and clear definition for clinical practice. In their work, they also acknowledge the variability of the crisis definition over the course of the literature, which has also been represented in this literature review. Tobitt & Kamboj highlight that there is some contention in the literature about whether Crisis Theory is in fact applicable in the context of community mental health practice, stressing the lack of clarity in the current definitions offered by the literature.

Tobitt & Kamboj investigated CRHT team workers’ understandings of the form and nature of the concept of crisis using qualitative techniques from framework analysis. They assessed interviews from 39 CRHT workers who were representative of the multidisciplinary skill mix of the four CRHT teams who participated. The participants completed semi-structured interviews which covered two broad areas of the crisis concept and CRHT work which were developed by the authors and were based on the literature. They found participants to hold a consensus about the concept of crisis. There were two main process areas identified through the analysis which linked to 1) identifying crisis and 2) understanding crisis. The continuum of crisis was shown to run from clearly ‘CRHT crisis’ to clearly ‘not CRHT crisis’. This continuum of crisis functioned under the influence of a number of modifying factors that were identified as disqualifying individuals from benefitting from CRHT intervention.

The outcomes of the research separated out crisis theory into theory that supported the understanding of the crisis phenomenon as it presents itself in individuals experiencing acute mental health crisis and theories that identified the origins or precipitators of crisis. Of particular interest to this research is the theory they developed to understand the crisis phenomenon. This theory contains information on items that are particularly helpful for the purposes of developing an assessment tool.

Three themes were identified as representing a present crisis state:

1) Functional Disruption – describing the temporary loss of functioning. This is the individual's ability to function, to cope, to care for themself and to have a sense of mental control.

2) Risk of Harm - this was described with a primary focus on risk of harm to self but risk of harm to others and risk of harm from others was also expressed by participants.

3) Additional Support Needed – this was characterised by failure of previous support and the requirement for further support to be put in place to support the individual to function in the community.

They found a further two themes that were described as often present 1) Extreme Mental Distress and 2) Otherwise Hospitalise. The first subtheme describes an increase in mental distress. However, an area requiring further clarity appeared to centre around whether or not this should be linked to previously existing mental health difficulties. 'Otherwise hospitalise' was surprisingly described by only a minority of the participants who expressed the possible necessity of hospital admission as a particularly pertinent theme for assessing crisis. Understanding this in the context of how and why CRHTs were developed, to act as gate keepers to inpatient admission, it may have been assumed that this theme would be considered important. It could be argued that this understanding is so implicit in terms of the functioning of the teams that this was not vocalised. Three themes were identified as always absent including 1) Referrer in crisis – this relates to inappropriate referrals which are the result of the referrer themselves becoming stuck or struggling to meet the needs of the patient. In this situation, it would suggest that the anxiety lies with the referrer rather than the patient themselves and therefore is considered inappropriate for CRHT intervention, 2) Longstanding 'crisis' - crisis was described as only a temporary disruption to functioning by the participants of this research and therefore by the very nature of crisis described it would not be a long standing condition and 3) Crisis level too mild – whereby the presentation of the patient was believed to be too mild to represent an acute mental health crisis.

Tobitt & Kamboj also identified theories of crisis in terms of the development of the phenomenon or crisis state. They describe a theme of *differential vulnerability* to help understand and describe why two individuals subjected to the same set of conditions may not both enter into the acute mental health crisis state. This was attributed to variability in the individual appraisal styles and therefore interpretation and perception of situations. There was also a theme around crisis presenting itself as a *time of opportunity* for individuals who took the experience of crisis as a chance to re-evaluate their lives, to make changes, learn and grow. *Readiness to act* was the third theme identified for theories of crisis. This mirrored the previous theme of crisis being a time for opportunity. For some individuals, a period of crisis in their lives is particularly motivating for making positive changes and work towards a brighter future.

The Tobitt and Kamboj study acknowledged the variability in the conceptual understanding of crisis and the need for further coherence. Developing a measure out of the emerging evidence for crisis theory and current working policy will support this coherence of understanding and therefore assessment. They did identify the broad and diverse needs presented by individuals presenting in crisis which will demand intervention in a variety of areas dependent on the individual need. Ideas around appropriateness of specific service user presentations was discussed with this idea that not all individuals will be appropriate for CRHT intervention even though their presentation may be very similar to others presenting with similar levels of distress or presenting ‘symptoms’.

Tobitt and Kamboj recognised that they did not extend their research to include individuals who had experienced crisis and saw this as a limitation in their work. This research was extended to patients but it is recognised that the carer perspective would also be important to capture. However, conceptually this may have quite a different orientation and may be worthy of separate investigation.

1.0.3 Implications of Crisis Theory on Crisis Treatment Approaches

The crisis theory outlined above implies that the experience of crisis is an acute, short term state of disequilibrium which results in the individual becoming emotionally and psychologically open and vulnerable. This may be a period of time where the individual learns and grows from the experience, resulting in an improved level of functioning post crisis compared to the pre-crisis state. Alternatively it could result in the individual struggling to resolve their crisis in a helpful and healthy fashion that ultimately leads that person to function at a lower level post crisis compared to their pre-crisis state. There may be a number of factors that influence the outcome of crisis for an individual, not least the support and treatment they receive from social and professional networks.

The experience of crisis opens a person up psychologically, making them more emotionally accessible (Rapoport, 1967) and amenable to suggestion and change (Golan, 1978). This is because the experience of crisis is a period of disequilibrium associated with stress. A person will experience associated anxiety as a result of this disequilibrium which causes a level of discomfort that acts as an impetus for change due to the person being unable to tolerate such high levels of discomfort or distress for long periods of time (Janosik, 1994). During a state of imbalance, the individual will strive to regain homeostasis or equilibrium, making the individual more open to intervention at this time. This means that the person is more motivated, which suggests there are significant treatment opportunities during this period to support the person to make a helpful crisis resolution. A focused intervention at this time of crisis may be more effective and have a greater impact than waiting for the person to be more stable but less ‘psychologically open’. A comprehensive assessment of the crisis is required in order to understand, evaluate and monitor the crisis state as that person works towards resolution.

1.0.4 Associated Theories Underpinning Crisis

A number of other theories underpin our current understanding of acute mental health crisis. These theories contribute additional substance and content to the concept or construct itself and so contribute to crisis measure development. A brief summary of these theories and how they link in with crisis theory is outlined below.

1.0.4.1 Homeostasis, Equilibrium and Crisis

Cannon (1932) developed the concept of homeostasis which presents the idea that some organisms i.e. endotherms, strive to maintain an internal constant state despite external changes. This requires energy to resist change and maintain optimal internal conditions and therefore does not occur by chance but is the result of organised self-government. Caplan (1961) related this concept to crisis theory in which he refers to psychological homeostasis whereby an individual strives to maintain psychological balance despite external changes. Cannon and Caplan's theories fit well with the research of Keyes & Ryff (2000) who found that individuals who had experienced change, positive or negative or both, found change unsettling and distressing. Therefore a movement away from homeostasis, even when positive, can cause discomfort and the individual will strive to return to the balance of homeostasis where they feel safest.

Taplin (1971) critiqued this definition as implying that the human psyche is mechanical in nature with no involvement of personality, feelings, ideas, skills etc. Homeostasis suggests a return to the pre-crisis state or normal balanced 'resting state' without judging how adaptive or maladaptive it is, therefore not allowing for learning or growth. The theory of equilibrium however, suggests that a system will strive to reach an equal balance, an equality of distribution and therefore the focus is on harmony with the environment. In contrast, homeostasis suggests the system strives to maintain levels independent of its environment. A state of imbalance, whether the result of a move away from equilibrium or homeostasis, takes the individual out of their comfort zone and results in feelings of being out of control and unable to cope. Intense feelings of helplessness, confusion, anxiety, shock, and anger are experienced as a result at these times (Golan, 1978).

There appears to be room for aspects of both theories in understanding crisis. As recognised in the work of Keyes & Ryff (2000), individuals feel uncomfortable in the context of change and will strive to maintain stability. However, if these attempts to maintain stability fail, the individual is forced to adapt and change to meet the new psychological demands of the environment. This may reflect the 'maximal resourcefulness' phase identified by Hobbs (1984) on the crisis curve (Section 1.0.2.3). The level of ability to function within the new context (the result of change) would depend on the success or failure of the individual to abandon the desire to achieve homeostasis and their ability to find a new balance, a new equilibrium in their new circumstances.

Both theories of homeostasis and equilibrium suggest that one of the underpinning concepts for crisis theory is this idea of change away from the person's normal functioning depending on what 'normal' functioning was in their pre-crisis state. In addition, there is also this idea about striving to return to the pre-crisis state which may present itself as resistance to move to a more positive position. What is very clear is that change results in the person feeling uncomfortable at best and acutely distressed at worse. One of the mediating factors is the individual's ability to be flexible and adaptable to change in order to reach a new balance or equilibrium. These theoretical points will be useful for considering the item pool developed through this research in terms of understanding the resilience or buffering factors and vulnerabilities that may make an individual more or less susceptible to experiencing crisis.

1.0.4.2 Maslow's Hierarchy of Needs

Maslow (1943) suggested that as human beings there are certain *needs* that we try to meet. He suggested that these needs exist in a hierarchy often illustrated in a pyramid form, whereby lower level needs along the baseline have to be met before attention can be focused on achieving higher level needs. The first and most basic level on the hierarchy is focused on meeting essential needs such as food, shelter, water and feelings of safety. Humans strive toward the top level of the hierarchy which is self-actualisation whereby the individual is able to recognise what their full potential is and make efforts to reach that potential. However, to master this top level need, all the previous levels of esteem, love and belonging, safety and physiological (basic) needs must be mastered first. At the base of Maslow's pyramid are the foundations for reaching individual potential. Without these in place the pyramid crumbles. This hierarchy clearly outlines the link between the necessity of basic physiological needs that provide the foundations for achieving and meeting higher psychological needs. The treatment of crisis will involve ensuring that the more basic needs are met before the higher order needs are facilitated.

Crisis theory suggests that an individual experiencing acute mental health crisis functions at a lower level compared to where that person would have been on the hierarchy prior to crisis. It is important to appreciate that recovery from crisis or returning to pre-crisis functioning will depend on the level of mastery prior to crisis. When an individual not only recovers from crisis but grows and learns from the experience, that person may progress to a higher level on the hierarchy. Where an individual is unsuccessful in attempts to positively resolve crisis, it may be expected that the person will continue to function at a lower level than the pre-crisis state. The crisis theory outlined in section 1.0.2 explains that one of the most prominent symptoms of crisis is the significant breakdown in ability to cope. It may therefore be expected that a person in the acute phases of crisis may demonstrate an inability to tend to basic needs or everyday tasks, at the base of Maslow's hierarchy, and these may be helpful item indicators for a tool designed to measure crisis. Where the foundations are shown to falter, it can be expected that the remaining levels on the hierarchy will become unstable.

It should be recognised that Maslow developed the hierarchy of needs from the perspective of an individualistic society whereby the development of self is at the pinnacle of the pyramid. This was criticised by Cianci & Gambrel (2003) who recognised that the pinnacle of the hierarchy for individuals from collectivist societies would be the need for acceptance and community, far outweighing the needs of self and self-development. The nature of the society would need to be taken into account in the application of this theory for assessing crisis.

1.0.4.3. Erikson's Theory of Life Stages

Negotiation of developmental crises is part of normal personality growth (Erikson, 1968). Erikson proposed that there are a number of psychosocial tasks that must be mastered as part of development. Each task or stage poses the threat of crisis due to the perceived threat of change. Where it is perceived that the task is insurmountable, successful completion of the next task will be hindered. There are two assumptions with this theory: 1) the world gets bigger as we go along, 2) failure is cumulative. The second of these assumptions is particularly pertinent to crisis theory in its suggestion that previous failure to successfully resolve crisis (in this context developmental crises), results in the individual being more vulnerable to experiencing crisis in the future. "The developmental process thus may involve a synthesizing of new experiences into an evolving self-perception and/or the accumulation of skills or strategies for instrumental or emotional response." (Turner & Avison, 1992; p.37). This understanding provides further support for the idea that an individual's history is the best predictor of their future and evidence of previous coping or lack of coping will be an important tool for the measurement of crisis. It also emphasises the idea that the resolution of the current crisis is critical if the individual is to develop a resilience rather than vulnerability to future crises. Erikson suggested that the experience of crisis could be a situation where growth or harm is experienced and this concept is also found in the theory of acute mental health crisis. This provides further theoretical support for the importance of timely and appropriate intervention to positively resolve crisis which would require comprehensive assessment to guide treatment decision making. The different perspectives around ability to manage stress based on previous successful/unsuccessful crisis resolution may best be explained by coping theory outlined in the section below (section 1.0.4.4).

1.0.4.4. Coping, Cognitive Appraisal and Crisis

Coping strategies support a person to manage and resolve difficulties, to deal with pressures and manage stress. Folkaman and Lazarus (1984) defined coping as "constantly changing cognitive and behavioural efforts to manage specific external and/or internal demands that are appraised as taxing or exceeding the resources of the person" (p.141). At a time of crisis it is acknowledged that previously implemented helpful coping mechanisms fail and the individual is left feeling unable to manage (Aguilera 1998; Roberts, 2005). Previously helpful coping strategies may lose their effectiveness in new or different contexts and the individual may move towards less helpful coping approaches in a desperate bid to resolve their difficulties and move away from the distress they are experiencing.

A stressor is a factor perceived as a threat to well-being. However, it is not the stressor itself that poses the threat to wellbeing but the individual's cognitive appraisal of that stressor. The individual cognitively appraises how well or not they are resourced to manage the stressor and the outcome of this appraisal is an overall perception of their ability to cope with it (Folkaman & Lazarus, 1984). "Since the 1960s there has been growing recognition that while stress is an inevitable aspect of the human condition, it is coping that makes the big difference in adaptational outcome." (Lazarus & Folkman, 1984, p.6). The perception of how stressful an event is by the individual is based on their perception of how well equipped that person is to manage it. "We approach this question through the examination of two critical processes that mediate the person-environment relationship: cognitive appraisal and coping" (p.19). The individual's interpretation of the situation, e.g. whether it is a 'good' or 'bad' situation is therefore based on the person's cognitive appraisal and subsequent perception of that stressor and their ability to successfully cope with it.

Events perceived as threatening, uncontrollable, or unpredictable are more likely to result in mental and physical illness (Thoits, 1995). Again, it is this idea of perception that is particularly important – it is not whether the event is or is not in fact threatening, uncontrollable, or unpredictable but whether the individual perceives it that way. An individual who perceives difficult situations as manageable and resolvable is much more likely to cope. A crisis occurs when an individual perceives an inability to cope and their normal coping strategies are shown to be ineffective, unhelpful or at worst detrimental.

According to Caplan (1964), there are seven characteristics for effective coping behaviour in crisis. These are:

1. Actively exploring reality issues and searching for information.
2. Freely expressing both positive and negative feelings and tolerating frustration.
3. Actively invoking help from others.
4. Breaking problems into manageable bits and working through them one at a time.
5. Being aware of fatigue and pacing coping efforts while maintaining control in as many areas of functioning as possible.
6. Mastering feelings where possible, being flexible and willing to change.
7. Trusting in oneself and others and having a basic optimism about the outcome.

These characteristics relate to internal coping mechanisms such as problem solving and ability to tolerate distress as well as the ability to utilise external resources to cope, with the aim of managing and adapting to change. These characteristics will be revisited in the development of the crisis measure (Chapter 2).

In terms of what cognitive appraisal and coping theory contribute to the understanding of the crisis construct, it is clear that an individual's ability to process, appraise and make sense of information in a healthy way will have a significant impact on the ability to cope and manage the situation. Therefore, it may be expected that an individual in crisis may experience a significant breakdown in their ability to cope and may display behaviours suggesting that their cognitive appraisal processes are no longer helpful at that time. This theory becomes useful in Chapter 3 when the identified subscales for the measure are appraised (Chapter 3, sec 3.9).

1.0.4.5. Ego Strength

Although borrowed from psychodynamically influenced literature, the concept of ego strength is useful for understanding one of the possible triggers to the crisis state. As outlined by Kanel (1998), "ego strength refers to the ability to understand the world realistically and act on it to get one's needs and wishes met. Many times a crisis worker will be called on to be the client's ego strength temporarily (for example when a person is psychotic or vegetatively depressed) until the client can take over for him or herself." (p.6). In line with Erikson's (1963) work, it is suggested that if a person has successfully resolved previous crises their ego strength will be greater and they will have a greater ability to successfully resolve crises in the future. If a person has not been able to successfully resolve previous crisis it would be expected that the person would have lower ego strength and would be more likely to fail at attempts to successfully resolve crises in the future. This would make the person more vulnerable to experiencing further crises as a consequence. Freud (1923; cited in Dufresne, 2011) referred to the idea of psychic energy, suggesting that it is finite with only a limited amount existing for each of us. This offers useful explanatory power for describing the experience of crisis to patients and staff alike. It strongly suggests that an individual will be more vulnerable to crisis at different times e.g. when a person is dealing with a large number of stressful situations at the same time they will be more vulnerable to feeling overwhelmed and as a consequence to experiencing crisis. This links in with the concept of ego strength suggesting that at different times we will have differing levels of ability to understand the world in a realistic way, respond to it appropriately and therefore act on it to get our needs and wishes met. This, again, relates to an individual's ability to cognitively process information with the suggestion that the crisis state would hinder this process resulting in a person less able to make helpful decisions and more dependent on others to support with this.

1.0.5 A Working Crisis Definition – Summarising the Crisis Theory.

To this point, this chapter has comprehensively reviewed the current theoretical understanding of crisis and some of the underpinning theories that support this understanding. It is important now to develop a succinct working understanding of crisis on which to base this research.

In summary, the theory of acute mental health crisis started over 100 years ago with more focused work starting with Lindermann in 1944. Lindermann's work focused on grief reactions but this was later expanded by Caplan (1961) to include all situational crises. In the last 10 years our understanding of acute mental health crisis has been pushed forward and clarified through the work of pioneers in the field such as Roberts (2002). The historical theoretical development, recent continued development and underpinnings from a number of different theoretical contributors has resulted in this current understanding and definition of crisis which blends several theories together to describe crisis as:

> All individuals have the potential to experience an acute mental health crisis but they will have differential vulnerability. A crisis state is triggered by a single or series of events that are perceived by the individual to be a threat to their life goals or values. The individual will make failed attempts to utilise their own known coping strategies and will strive to utilise new coping strategies that may be helpful or unhelpful, possibly leading to harm in a desperate bid to alleviate distress. The failure of coping results in a period of psychological imbalance that is distressing and perceived as unmanageable by the individual. The crisis state is demonstrated by functional disruption, risk of harm to self or others, extreme distress and the need for additional support. The crisis state is time

limited to between 4-6 weeks and in this period the person is more amenable to suggestion and change and therefore more motivated to take on intervention. The outcome of the crisis state can be either positive or negative but the aim of crisis intervention is to take advantage of the possible opportunities presented by the crisis state to support the person toward the best possible resolution.

1.0.6 Models of Crisis Intervention

This research is focused on the development of a tool to measure crisis which will accurately support and inform treatment decisions made by those working in crisis resolution teams. It is helpful at this point to look to the literature for existing models of crisis intervention which informed the development of this crisis measure that ultimately aimed to support crisis teams in their decisions regarding intervention. A number of crisis intervention models currently exist and have been outlined in this section.

Crisis intervention is the process of interceding into a person's crisis experience, offering support to alleviate the pressure and subsequently guide the person towards accessing internal resources and independence with the aim of successfully resolving the crisis. The role of the crisis worker, family member or friend, is to support the individual through the crisis towards a more stable state, guiding toward regaining independence and moving away from dependence. If crisis intervention is to be helpful, more importantly not detrimental, it is essential that intervention approaches are developed from a sound theoretical and preferably evidence based foundation. This was eloquently expressed by Hoff (2001) who stated that: "Without a sound theoretical base and established techniques, there is little to distinguish crisis intervention from intuitive first aid." (p.66).

Aguilera's (1998) Problem-Solving Approach (Figure 1.2) to crisis intervention is based on the concept that an individual is constantly faced with problems that need to be solved as outlined by Caplan (1961). When problems are successfully negotiated the individual will maintain equilibrium. However, where the problem is not successfully solved this can result in a state of disequilibrium which may result in crisis. This can also occur when problem resolution takes an unusually large amount of energy, concentration, and resources. This links in with the underpinning theory of ego strength and psychic energy outlined in section 1.0.4.5.

Figure 1.2 Aguilera's (1998) Problem-Solving Approach

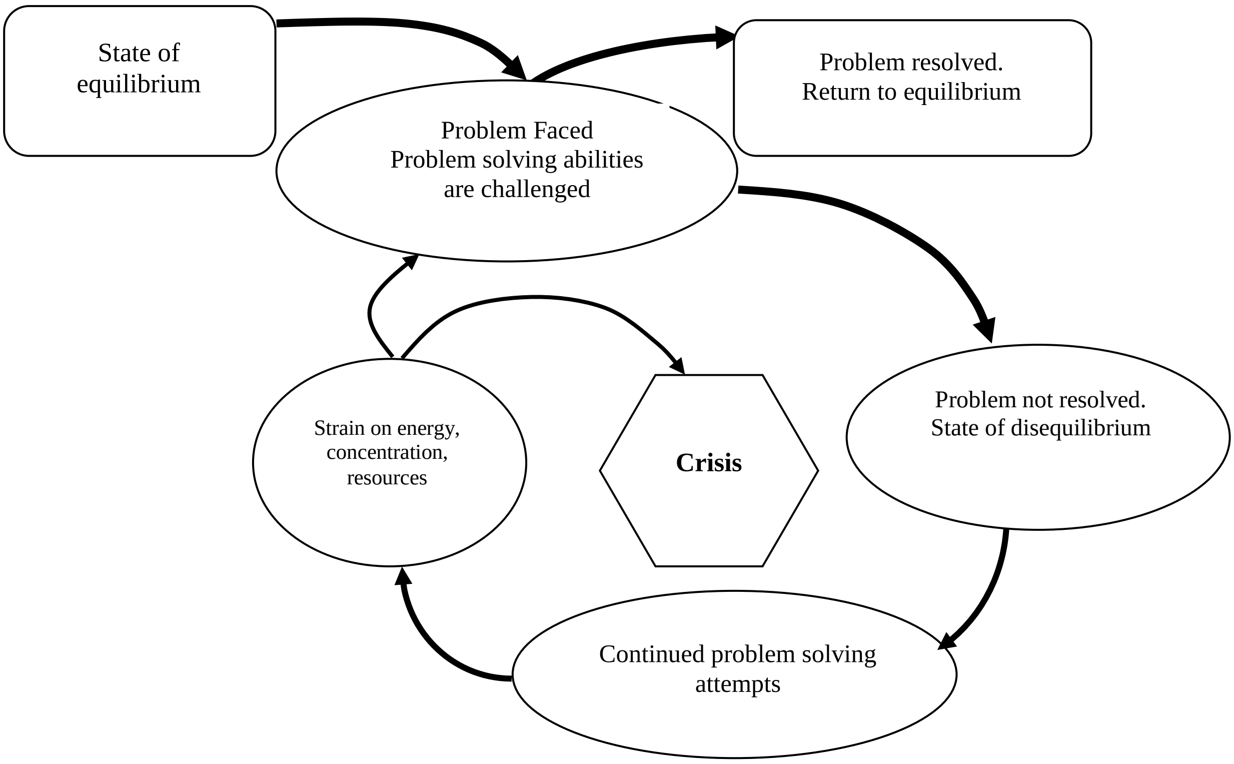


Figure 1.2: Aguilera's (1998, p.26) Problem-Solving Approach to crisis has been depicted here as a model with the conceptualisation of a spiral. Where effective coping breaks down, the individual is taken out of their natural equilibrium state, resulting in a crisis spiral that leads to a state of acute crisis.

Aguilera's (1998) approach to crisis resolution aims to support the individual to problem solve in a more healthy, constructive and successful way. The goal of this approach is to re-establish the pre-crisis level of equilibrium and functioning or to work towards improvement. Aguilera proposed a four step problem solving approach to crisis intervention. 1) assessing the individual and the problem, 2) planning a therapeutic intervention with a focus on problem solving skills directly related to the problem identified in step one, 3) the intervention itself and 4) finally anticipatory planning for future potential problems. Aguilera divided the intervention phase into four further sub-sections a. helping the individual to gain an intellectual understanding of the crisis, b. helping the individual bring into the open their present feelings to which they may not have access, c. exploration of coping mechanisms, d. reopening the social world. It is not surprising to see that the very first step in this intervention model is the assessment of the individual and their problem. In fact, it is common across the crisis models (see below), that assessment is an integral aspect of the process of crisis intervention if successful crisis resolution is to be achieved.

Kanel (1998) proposed the ABC model of crisis intervention roughly based on Jones' (1968) A-B-C method of crisis management, which made up of three stages. The first stage (A) focuses on developing and maintaining contact with the individual experiencing crisis, this is in line with the literature showing the therapeutic relationship to be one of the most important ingredients for the success of a therapeutic intervention (Messer & Wampold, 2002). The second stage (B) looks at identifying the problem (assessment) and subsequently the therapeutic interaction. In practice, the therapeutic relationship and exploration of the crisis precipitating problem would probably occur in parallel. The assessment stage attempts to understand the client's perspective of the problem, therefore allowing the development of an intervention to support a healthier problem or crisis perspective. The therapeutic interaction can only be developed following a thorough assessment to identify its focus. The final stage (C) looks at identifying current helpful and unhelpful coping strategies with the aim of maintaining the helpful ones, reducing the unhelpful ones and adding further helpful coping strategies.

Roberts (2005) developed the seven-stage crisis intervention model (Figure 1.3). Developed from the critical stages identified in recovery journey of crisis, it integrates the stages proposed in previous models such as Aguilera's (1998) and Kanel's (1998) above. The seven stages are outlined below:

Figure 1.3 Roberts' Seven Stage Crisis Intervention Model

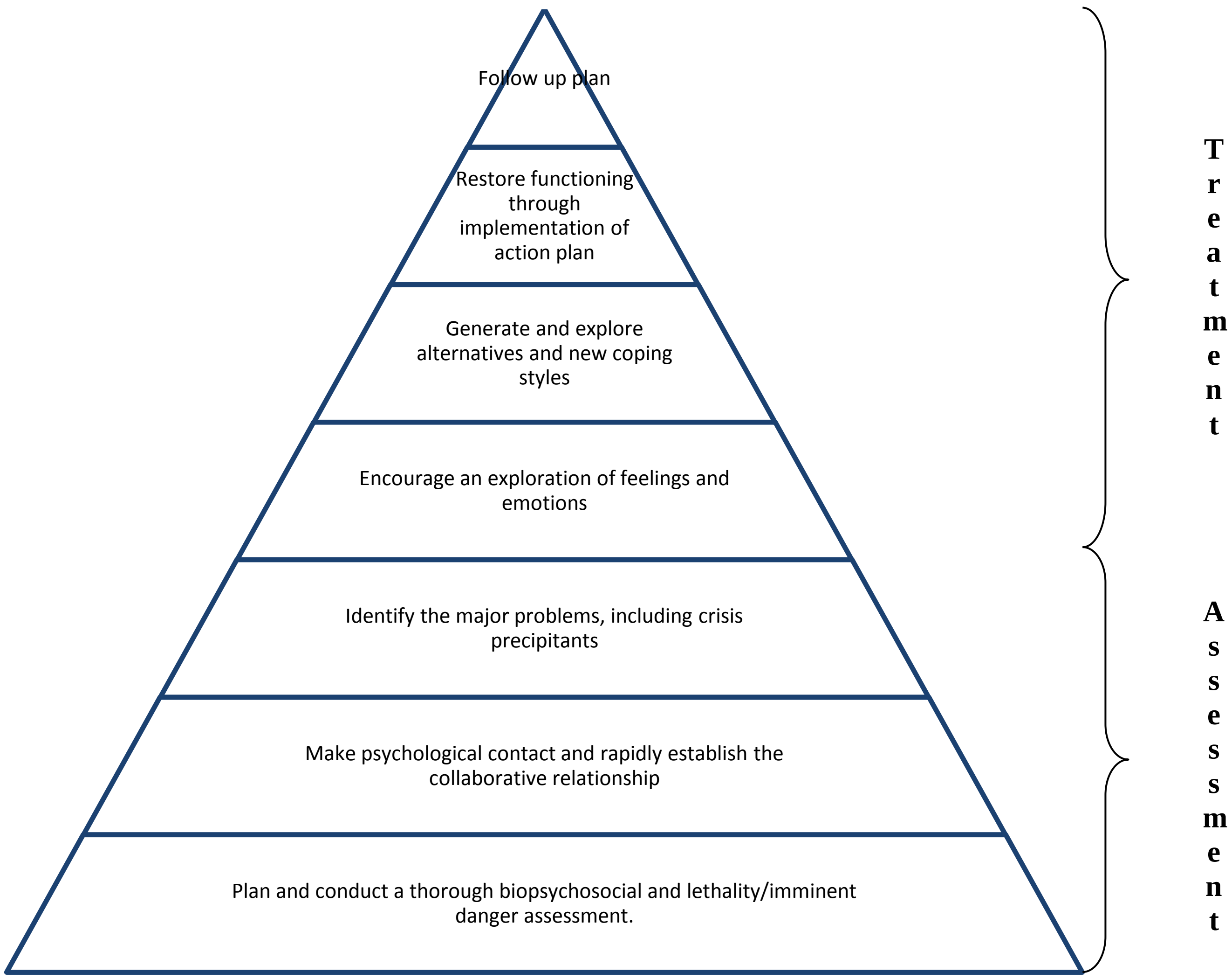


Figure 1.3 Roberts' (2005, p.20) Seven Stage Crisis Intervention Model outlining the stages of assessment and intervention recommended for the successful treatment of crisis. The first three stages, forming the foundation of successful crisis intervention, relate to comprehensive crisis assessment.

Assessment provides the foundation upon which treatment is based and crisis resolved, contributing to the first three stages of Roberts (2005) model. The unpredictable and changeable nature of crisis means continuous assessment of a patient's needs is critical to ensure that the care plan continues to be relevant and helpful. The treatment model presented by Roberts is currently the most comprehensive available. Roberts emphasises here the crucial nature of assessment as the foundation of all further work by placing it as the base of the treatment pyramid.

The Sainsbury Centre for Mental Health published the Crisis Resolution and Home Treatment practical guide in 2006 (Ford & McGlynn, 2006). This looked at aspects of CRHT teams including the role of values in crisis, risk management and working through a crisis. A basic outline of the work of CRHT teams by McGlynn and Flowers (2006) indicated a four-stage model which is grounded in assessment as the first stage. The first stage is described as accepting the referral and assessing the crisis. This requires identification of contributory factors, the modifying factors and the strengths which might help the individual to resolve the crisis. They also emphasise the necessity of identifying who is involved in the crisis in both helpful and unhelpful ways. The second stage focuses on planning by developing strategies for managing the crisis preferably in the community setting. It specifically looks at a support network approach whereby the skills of professionals, the individual and others who can support the person are drawn upon. This again emphasises the importance of the community network in collaboration with the mental health professionals and should therefore significantly contribute to the assessment of crisis. The third stage looks to implement the intervention with the fourth stage looking to identify when the crisis has been resolved and professional support can be reduced. It is clear when looking at this four stage process that monitoring and measurement are crucial in order to complete a comprehensive assessment and set a baseline against which progress and crisis resolution can be compared. In the description of this model it is clear that all decisions about a patient's care and treatment are based on the initial assessment, followed by the continuous observation and assessment by staff. Without a standardised, valid and reliable assessment tool this process will not be able to provide meaningful outcomes.

1.0.7 Current Treatment of Crisis in the UK

The community treatment model suggested by Lindemann (1944; 1956) was not adopted in the UK until over 55 years later when the NHS Plan (Department of Health) was published in 2000. This outlined the requirement for 335 Crisis Resolution and Home Treatment (CRHT) teams covering the whole of England. CRHTs are specialist multidisciplinary teams that operate 24 hours a day, 7 days a week offering emergency assessment and intensive home treatment as an alternative to acute hospital admission. The current working CRHT model developed over the last 10 years has its roots in America and Australian crisis team models (Johnson & Thornicroft, 2008). Research to date has shown that CRHTs have significantly reduced admission rates (Glover, Arts & Babu, 2006; Johnson et al, 2005; Keown et al, 2007), reduced bed occupancy (Keown et al, 2007) and reduced assessments under the Mental Health Act (Dunn, 2001a,b).

The time scale outlined by Lindemann (1944) and the call for short, focused, community intervention is now embedded in the operating procedures of CRHT teams in the UK today (DH, 2001a). In the UK, the Department of Health (2001) CRHTs were set up to support adults between the ages of 16 to 65, experiencing acute mental health crisis and ‘‘with an acute psychiatric crisis of such severity that, without the involvement of [CR/HT], hospitalisation would be necessary’’ (p. 11). The CRHT teams engage with patients for up to a maximum of 6 weeks, in line with the work of Caplan (1964). This does not imply that the crisis has been completely resolved but that the individual has moved out of the acute crisis state as outlined by Roberts (2005). It is the aim of the CRHT to move the individual out of the state of crisis towards resolution of the problem. However, longer term presenting difficulties are treated as enduring mental health problems by other services such as Community Mental Health Teams (CMHTs).

In December 2007 the National Audit Office published ‘Helping people through mental health crisis: The role of CRHT Services which outlined the evidence base that indicates where CRHTs are used appropriately and safely results in a number of clinical benefits which ultimately leads to improved patient satisfaction. They also explored the impact this has had on reducing stigma and social exclusion which is often faced by individuals experiencing acute mental health difficulties.

This work suggested that to improve the appropriate use and safety of CRHTs, a standardised, valid and reliable outcome measure was required. The need for clinically credible outcome measures (e.g. tools, psychometric measures) to be used by all NHS services was also outlined in the Liberating the NHS white paper (2011) as a means to assess and monitor clinically meaningful change and progress.

1.0.8 The Role of Crisis Assessment in CRHTs

Effective crisis intervention is dependent on accurate assessment that directly translates into focusing treatment where it is needed (Myer & Conte, 2006). Crisis assessment should support the CRHT team to decide firstly if a crisis state is present and the nature of the crisis, secondly if the patient is suitable for CRHT intervention e.g. can the crisis be safely managed at home, thirdly to inform and support care plans that should be based on short focused treatment interventions and finally to track recovery and inform potential referrals to other services (Hoult & Cotton, 2008).

Standardised assessment approaches may support mental health workers to maintain a comprehensive, standardised and consistent approach and ensure that a baseline of assessment is achieved for each patient that will support professionals and patients to track change. In acknowledgement of the many advantages a standardised assessment measure would offer, Roberts & Lewis aptly stated “What is remarkable, however, is the lack of empirically based and standardized crisis assessment protocols, including instruments with strong psychometrics that are available to practitioners.” (2001,p.18). This challenge has still not been met and the problem remains.

Assessment for crisis should ensure that the level of risk and need by each person is understood therefore supporting appropriate and focused treatment to be delivered. The nature of crisis is that change is continuous and rapid. Therefore it is essential that crisis team workers continuously assess a patient’s presentation to ensure that the treatment approach being delivered is the one most suited to the individual’s need at that time. This supports the call for a standardised assessment tool to allow continuous standardised assessment to be completed.

Assessment in crisis intervention has different goals, processes, relationship to treatment, and types of information gathered when compared to other teams (Myer & Conte, 2006). The very nature of a crisis has challenged the development and implementation of lengthy assessment tools and procedures because the process for crisis intervention must be fast and continuous. Crisis assessment occurs immediately and very often so too does the crisis intervention. In comparison, current assessment process by professionals in Community Mental Health Teams (CMHT) is carried out over a number of weeks with the comprehensive collection of information regarding the patient's history, interviews with the patient and relevant others and completion and analysis of assessment tools.

An assessment structure is necessary to support CRHT assessors in understanding the patient's presentation, to help the patient understand their own presentation, to standardise the assessment procedure to ensure quality, to enable mapping of patient progress, to demonstrate outcomes and to show treatment effectiveness. This outlines the important role that crisis assessment should play in the work of CRHT teams if a comprehensive understanding of a patient's crisis is to be obtained and appropriate treatment intervention delivered in a timely manner.

1.0.9 Approaches to Crisis Assessment and Crisis Measurement

Approaches to crisis assessment have been outlined by a number of authors, practitioners and researchers and a small number of assessment tools have been developed for this purpose. The assessment tools developed to date have been in the form of either self-report inventories or clinical ratings following a guideline or protocol.

One of the primary purposes of crisis assessment is to identify if the patient is in fact in a state of crisis, suffering with some other mental health problem or does not demonstrate mental health difficulties at all. Golan (1969) placed emphasis on identifying whether or not the patient is in an active crisis state at the time of the initial assessment and outlined two questions as crucial for assessing this: When was the precipitating event?; How long has the individual been trying to cope with this on their own?. To identify a crisis state, Golan outlines four components devised from the work of Caplan, (1964), Klein & Lindemann (1961) and Sifneos (1960): 1) the hazardous event, 2) the vulnerable state, 3) the precipitating factor, and 4) the state of active crisis (disequilibrium). Golan stresses that this is not a rigid structure but simply provides a framework or guidance for identifying the active crisis state. Golan proposed a model for an intake interview for crisis work, stressing the importance of working quickly and comprehensively at the time of crisis, recognising that at the time of active crisis, the patient is most open to intervention and change. The majority of patients seeking crisis intervention are either looking for symptom relief or relief from external pressure, both of which make the person, who may normally be quite rigid, more open to making changes.

Once it has been shown that the patient is in an active crisis state the team are in a position to carry out a comprehensive assessment of the crisis. This is important in order that the intervention phase of the treatment is well informed and will support the individual towards positive change and resolution. It is also possible to combine these steps (step to identify crisis and step to identify state of the crisis presentation) by placing crisis on a continuous spectrum running from *no crisis* to *acute/severe crisis*.

The Crisis Rating Scale was developed by Bengelsdorf, Levy, Emerson, and Barile (1984). This required the clinician to rate the three dimensions of dangerousness, support systems and ability to cooperate on a 5 point Likert scale. The limitations of this scale relate to content coverage of the crisis construct. Content coverage is the extent to which a set of variables are able to fully cover and describe all of the aspects of the construct of interest. It is questionable whether or not the three items outlined in the Crisis Rating Scale are sufficient to provide content coverage of a construct as complex as crisis. In addition, the analysis of the rating scale to ensure fundamental measurement and analysis of the scale's unidimensionality were not completed. These three aspects of measurement are crucial if it is to be inferred that the measure's outcomes provide a meaningful representation of the crisis state. Rating scale functioning and unidimensionality are outlined in more detail in section 1.2.1.2. Although this was a helpful first attempt, the construct of crisis was not thoroughly investigated in developing these dimensions and the scale itself not assessed for validity and reliability. Validity and reliability are integral to providing evidence of the quality of a measure and therefore are vital if a measure is to be implemented with any confidence. This research aims to address the weaknesses demonstrated in previous attempts to develop an outcome tool for crisis assessment and will utilise techniques that are specifically aimed at determining and refining the psychometric properties of assessment tools.

Perlmutter & Jones (1985) outlined the Psychiatric Emergency Service Interview (PES Interview) for working with families who present at the emergency department. The outline shows that the assessment opens with a focus on engagement, reducing anxiety, validation and identifying the specific problems of the crisis. Similar to Kanel's (1998) ABC model (section 1.0.2.3), the focus here is on developing a healthy therapeutic relationship. Crisis assessment is an integral part of any crisis intervention but underpinning any successful intervention is a healthy therapeutic relationship based on openness and trust. Following the initial step of forming a relationship and identifying the crisis problem, Perlmutter & Jones recommend completing a traditional assessment of the patient including an assessment of organic root causes of changes in behaviour, assessment of bizarre behaviour, ability to care for self and any potential for harm to self and harm to others. Finally, they suggest a specific assessment of the social context including the social and family systems. It is interesting that the social context has been given a specific focus and links in with the underpinning theory of Maslow's (1943) hierarchy of needs (section 1.0.4.2) which will be specifically explored in more detail through the process of this research (Chapter 3, section 3.10). Some of the areas of assessment identified for this approach will be included in the final assessment tool developed through this research.

The importance of engagement and the development of a positive therapeutic relationship is highlighted here. The interesting viewpoint for this assessment approach is that it looks at the involvement of the family and the possible benefits and/or harm that they could pose. Family and supportive networks are described as being pivotal to decisions regarding admission versus home treatment and are also reflected in the findings of this research (Chapter 3, section 3.10.4).

Myer, Williams, Ottens and Schmidt (1992) developed the Triage Assessment Model which was later developed into the Triage Assessment System (TAS) (Myer & Conte 2006). As highlighted by Roberts (2005) this assesses affective (emotional), behavioural, and cognitive domains of individuals reacting to crisis events on a 10 point Likert scale of impairment. The benefit of this approach is that it is continuous and can be used to monitor patients' progress. They outline approaches for assessing the three domains and tackle some of the difficulties that may interfere with this process. The key to assessing the affective domain is to identify the principle emotion and, generally for individual's experiencing crisis, this will either be anger/hostility, anxiety/fear, and sadness/melancholy. The authors identified these emotional reactions from research on the primary emotions by the National Advisory Health Council (1995) & Plutchick (1980). Behavioural reactions were categorised as immobility, avoidance, and approach (Myer, Williams, Ottens, & Schimdt, 1992a). The cognitive domains are also grouped into three categories transgression, threat and loss. These categories were taken from the work of Lazarus on stress and ways of coping. Transgression is defined as meaning "demeaning offence against me and mine" (Lazarus, 1993, p.26) which relates to a violation of a law, duty or moral principle. Threat is the perception that something is going to happen in the future and loss is the perception that it has occurred in the past. They go on to outline the areas of individual's lives that are affected by the experience of crisis including physical, psychological/self-concept, social relationships and moral/spiritual and suggest that transgression, threat and loss can be perceived by patients in each of these life dimensions. The authors of this work looked to previous literature and theory as their method for defining the categories of this measure. However, similar to the criticisms of the other assessment outlines discussed in this section, the construct of crisis remained without direct qualitative investigation to unearth its true content. These authors simply looked at the work of other authors to develop their understanding of crisis to be rated on the measure. The validity of this approach for developing an understanding of crisis could be improved to enhance the representation of the concept or construct upon which a more comprehensive measure could be developed. In addition, the 10 point likert rating scale was not assessed to identify whether or not it truly represents interval level rating or if the measure is unidimensional.

Ryrie and colleagues (1997) outlined a 'zoning' system for managing case work for CMHTs. The zoning system is the same as the traffic light system adopted by the Bedford and Luton CRHT teams which are sampled in this research. Ryrie et al look at the concept of zoning from the perspective of supporting the CMHT team in managing their caseload more effectively to ensure that resources are targeted more appropriately depending on the level of client need. The assessment determines the zone that the client will go in and therefore the level of resource that will be received. The patient is assessed to determine how stable/unstable their mental health is and how well/unwell they are engaging with services. Patients who have unstable mental health and/or show a rapid decline in engagement are placed in the red category, the amber zone is for clients who are mentally unwell but who do not present any major risk factors, and the green zone contains clients who are stable and are being monitored or are receiving maintenance care. In addition, this system also had a black zone which represented clients who currently reside somewhere other than in their own home for example in hospital or prison. Ryrie et al recognised that the criterion for the zones is non-specific and relied on professional/clinical judgement but this flexibility of assessment was supported by the mental health professionals who acknowledge the advantages of this approach. The flexibility of an assessment tool to allow scope for intuition or 'gut instinct' may be helpful for an effective crisis assessment tool and so will not be ruled out in this research.

Robert's (2001) seven stage model of crisis intervention outlines the psychosocial and lethality assessment as the first stage of crisis intervention. The emphasis at this stage is on the crisis worker completing a swift and thorough biopsychosocial assessment (Roberts & Lewis, 2001). Roberts goes on to outline from the work of Eaton and Ertl (2000), that at a minimum the crisis assessment should cover a number of areas including 'the client's environmental supports and stressors, medical needs and medications, current use of drugs and alcohol, and internal and external coping methods and resources.' It is interesting that the assessment of the patient's environmental support systems and stressors should factor first on the list. The theory outlined suggests that a person experiencing crisis is unable to utilise successful coping strategies and therefore the individual will be dependent on others to cope for them in the initial stages. An individual with minimal support systems will show a greater need and dependency on the team than an individual supported by a close family. Again the mediating/protective factors are shown to be crucial to crisis assessment.

For Roberts the assessment stage of the crisis work is crucial for assessing support systems, emotional, cognitive and behavioural aspects of the crisis, lethality and danger as well as establishing a positive working rapport. If the rapport is not established and the client does not engage well with the services, working safely and positively with that patient will be more difficult if not impossible.

James & Gilliland (2001) proposed a six step model to crisis intervention. The first 3 steps have an overarching title of 'Listening' which includes 1) defining the problem, 2) ensuring client safety 3) providing support. The second set of three steps are summarised under the heading of 'Acting' and include 4) examining alternatives 5) making plans, 6) obtaining commitment. They use the concept of assessing clients for their state of equilibrium and mobility, where equilibrium is a state of mental or emotional stability and mobility is the state of the physical being whereby the person can autonomously change or cope in response to different conditions and is able to be flexible and adaptable to the social and physical world around them. In addition, they look at assessing the patient's psychobiological functioning, neurobiological changes, effect of legal and illegal drugs, support systems, coping mechanisms and assessing for suicide or homicidal intent.

Bonynge & Thurber (2008) attempted to develop clinical ratings for crisis assessment after recognising that "We have no way of classifying or describing crisis intervention episodes. There exists no reasonable explication of any construct we call crisis intervention." (p.304). They developed five variables from consensus of their crisis unit professionals – 1) danger to self, 2) danger to others, 3) functional decline, 4) confusion, and 5) depression. Exploratory factor analysis identified two factors that accounted for 43.1% of total variance with internal reliability of .68. Although a positive step forward in the direction of crisis measurement, it is questionable if content saturation of a construct as complex as acute mental health crisis can be obtained by the five items outlined for this scale. In addition, the unidimensionality of this scale or the two subscales to ensure that measurement is of one dimension or construct was not assessed. These criticisms appear to be common across the previous assessment outlines and will form a critical aspect of this research in addressing these research gaps.

1.1 Understanding Measurement

Based on the evidence from the initial literature review, it is clear that there continues to be an unmet need to develop a sound, reliable and valid crisis measure that can be used with confidence to inform the complex clinical judgements that have to be made by CRHT staff.

Measurement tools help to identify how much of a ‘thing’ or ‘matter’ is present. In psychology, measurement tools aim to measure psychological constructs such as acute mental health crisis. There are different types of measurement tools available with which to measure different types of matter. It is important that the right approach for measuring the ‘matter’ of crisis is used to ensure that the measure developed can provide outcomes that are meaningful and accurate. This section provides a summary of the evidence base from which the measurement development techniques for this research were chosen.

1.1.1 Fundamental, Derived and Conjoint Measurement

The basic aim of measurement is to understand ‘how much’ of something there is so that it can be used for comparison purposes when making judgements about what decisions to take. Comparison is only possible when measurement occurs on the same scale or metric and the units of measurement are all equal, for example the assumption that the distance between 1 and 2 is the same as the distance between 2 and 3 on the same scale. In the physical world, measurement can be a little easier to quantify by very nature of its accessibility whereby it is visible and tangible e.g. being able to directly measure the length of an object. When a thing can be directly measured e.g. weight or length, it is called fundamental measurement.

It starts to get more complicated when the thing of interest cannot be directly measured, for example temperature or density. Density is measured indirectly using mass and volume. Temperature is derived from the volume changes of mercury under the influence of the temperature. Both of these are called derived measurement approaches as they are derived indirectly through directly observing other related entities.

Conjoint measurement is based on the understanding that when a variable cannot be directly observed, it can be indirectly observed through the observation of changes in other related attributes (Berka, 1983). This approach is used in the development of psychometrics whereby the attribute itself is not measured but instead is measured through other observable attributes that are related to the attribute of interest.

Therefore, as measurement moves from fundamental to derived and finally to conjoint measurement, it is gradually becoming more and more removed from direct measurement. Psychological constructs such as crisis can only be measured through conjoint measurement techniques.

Constructs that cannot be directly observed or measured must be inferred and substances that have to be inferred are termed latent constructs (Atkinson & Lennox, 2006). Therefore when measuring crisis, a *latent construct* is being measured.

The more indirect and removed measurement becomes the more room there is for error to occur. This is immediately a challenge for developing a measure for crisis that can only be measured using conjoint measurement techniques. A further challenge in developing measures for complex constructs or attributes is that they often rely on the measurement of several different areas/skills/knowledge/symptoms to indicate where on the variable of interest the person exists. In this research, crisis is both complex and requires indirect measurement techniques that need to tap into a number of different areas to fully represent the concept or latent construct.

Although the measurement of latent constructs is a little more difficult than the measurement of directly observable physical world things or matter, the same principles of representative fundamental measurement still apply. The aim is to know how much of the latent construct is present, and if it is more or less than at a previous point in history or than another person. This measurement and comparison is only possible if the scale used is truly representative of the construct in question and the distances between scale points are equal.

In the past, the focus of measurement construction has largely been on the statistical techniques used to interpret the scale outcomes with less attention paid in general to the quality of the measure developed in terms of how the individual items and their rating scales are functioning. The scale outcomes are meaningless if care and attention have not been paid to ensuring that the measure has been designed to truly represent and therefore tap into the construct of interest. This will be discussed in more detail in the next section.

1.1.2 What's the Matter?

One of the most important first steps in measurement design is to ensure that a psychological measure is tapping into the construct of interest. To understand this concept it can be helpful to consider measurement in science as a comparison. In order to measure *matter* in science it is necessary to first know what the matter of interest is. Having a thorough understanding of the matter makes it easier to choose an appropriate measurement tool: for example, the techniques to measure a liquid will be very different to the techniques used to measure a gas or vapour.

In the physical sciences it is important to know the substance and structure of the matter being measured (IUPAC, 2006). The substance is the form of the matter, for example water is made up of the substances hydrogen and oxygen (H_20 has two hydrogen atoms connected to an oxygen atom). There may be a number of substances that make up the matter/construct of interest e.g., the water that comes through the tap is, in reality, made up of a number of different substances, therefore it is also important to know what the structure of the substance is. In chemistry this looks at how the atoms/molecules are arranged in relation to each other, which is analogous to looking at the structural relations between the items of an item pool in psychometrics. It is only by focusing in at the substance and structural levels of the matter that it is possible to understand what the matter is in order to give validity and reliability to any measurement attempts.

Understanding a psychological construct or attribute such as acute mental health crisis is similar. The first step is to fully understand what the substance of the crisis construct is, what ‘atoms’ (items) contribute to the crisis matter (construct). The second step is to understand how these items relate to each other and to try and identify if there is in fact one substance (a unidimensional construct) or a number of substances (constructs/subscales/multidimensional measure) that make up the construct of interest. One of the first steps in developing a measure for crisis is to thoroughly investigate the substance of the construct which is generally done through qualitative techniques to develop the item pool (Chapter 2). Following this the structure of the substance is investigated using factor analysis or Principal Component Analysis to identify the subscales. It is only when the substance and structure of the crisis construct have been identified that it can be measured in any meaningful way.

1.1.3 What are Psychological Measures?

Psychological measures are tools that obtain information believed to represent a psychological construct to which numbers are assigned according to specific rules (Stevens, 1946). Psychological measures are called psychometric when they place a psychological construct onto a metric which is a standard of measurement. Psychological measures are common for measuring intelligence, personality and other mental health constructs, made popular by the founding fathers of psychometric measurement such as Cattell (1886), Spearman (1906), Thurstone (1936) and Rasch (1960). Psychometric measures inform important and often life changing decisions, therefore the quality of the measure is paramount. Miller, McIntire & Lovler (2010) summarised the importance of well-designed psychometric measures stating that “Good tests facilitate high-quality decisions, and bad tests facilitate low quality decisions” (p.4). Placed in the context that the outcome of a psychometric measure may result in the difference between employing the right person, assigning a child to the right school or class or the difference between an individual receiving crucial timely support from mental health services or not, renders the design of the measure as critical to the development of a valid tool for crisis.

1.1.4 Why Measure?

The aim of this research is to develop a measure for crisis assessment so it is important to think about why measurement is important and whether this approach to assessment offers any advantages over previously utilised assessment techniques such as standard note taking or documentation.

1.1.4.1 Standardised and Comprehensive Assessment

Crisis is a complex presentation that is affected by biological, psychological and social factors. Due to the complexity of the crisis presentation it is impractical and unrealistic to expect an assessor to comprehensively document all of the contributing factors or variables by hand, particularly when it is important to note risk and protective factors as well as stable and unstable factors. Stable elements are important to monitor in order that deterioration and change over the course of treatment are shown by comparison to baseline measures. Without a thorough investigation into the construct of crisis there is a continuing potential to document unnecessary information at the same time as missing crucial nuggets essential to treatment decision making. The development of the item pool (Chapter 2) for a crisis assessment measure will provide an aide memoir and short hand for mental health professionals tasked with documenting a patient's presentation. This will be in addition to providing an indication of the level on the crisis construct.

Having a standardised set of items that are completed for all individuals assessed for crisis intervention ensures that every patient receives the same comprehensive assessment to support the delivery of a standard of quality. In addition, in recognition that there is no specialised training for mental health professionals working in CRHTs in the NHS at present, a standardised assessment measure would be a useful training tool to support newly qualified staff or staff new to the CRHT team to understand all of the areas necessary for assessment.

1.1.4.2 Comparison

A standardised set of items rated at different points throughout a patient's crisis journey supports the CRHT team to monitor change and progress. In addition, standardisation of a measurement tool allows the individual to be compared to different crisis states which have been identified through normative data. This will be a particularly valuable tool for CRHTs who could use this comparison as a thermometer for crisis, effectively providing an indication of crisis level.

1.1.4.3 Valid and Reliable Indications of Crisis Level

As outlined above, it is only when the substance and structure of the construct of interest has been identified that attempts at measurement can be made. The identification of the substance and structure of the construct provides evidence for the validity of the measure – that is, the construct being measured is the one of interest. Additionally, once data has been collected for the newly developed measurement tool, indications of reliability (the accuracy and stability) of the measure can be obtained through statistical analyses. The validity and reliability of measures are essential elements that provide an indication of quality and integrity for the measure. These will be discussed in more detail later (sections 1.2.3.2) in specific relation to the crisis measure being developed through this research.

1.1.5 Assumptions of Psychological Measurement

There are a number of assumptions that psychological measures should adhere to if they are to fall into the category of a 'good psychological test'. According to McIntire & Miller (2007) there are 6 main assumptions:

1. Psychological tests should measure what they say they are measuring. That is, they should have construct validity

2. Where the construct being measured remains stable over time, the outcome of the measure should also remain stable (good temporal reliability).

3. Test items should be constructed in a manner that will be interpreted in the same way by all test takers. Therefore test items should be well constructed with support and input from those who would be completing the measure (good construct validity and inter-rater reliability).

4. Individuals will be able to complete items accurately.

5. Individuals will be able to complete items honestly.

6. The test score is a representation of the true score with measurement error due to the test itself, the assessor, the assessed or the environment (Classical Test Theory).

One of the aims of this research is to develop a measure for crisis assessment that will meet all the assumptions outlined above by McIntire & Miller (2007). This will help to ensure that the resulting crisis measure falls within the parameters of the category for a ‘good psychological measure’. The framework above therefore acts as a checklist for this research and is used as part of the discussion of the developed measure from this research in Chapter 9 (section 9.1).

1.1.6 Developing Measures in Practice

1.1.6.1 Is there a Measurement Need?

The first step in developing a measure for crisis is to identify if there is a measure development need. If measurement is not required or if a valid and reliable measure of sound quality already exists for that measurement area then it would be a pointless exercise to develop a new measure. This is the first step on the Simms & Watson (2007) construct validity model (Figure 1.4). The literature review in the earlier sections of this chapter (section 1.0) clearly demonstrated that there continues to be a significant need for the development of a measurement tool to support the assessment of crisis by CRHTs and provides the rationale upon which this research is based (section 1.5).

1.1.6.2 Developing the Measure and Construct Validity

Construct validity is the process of developing and testing psychological theories and measures to ensure that they are truly representing the construct of interest accurately and fully (Loevinger, 1957). Construct validity is defined by Strauss & Smith (2009) as the "evaluation of the extent to which a measure assesses the construct it is deemed to measure". It accounts for all forms of validity including content, predictive, concurrent or empirical validity (Landy, 1986; Messick, 1995; Strauss & Smith, 2009). Messick (1995) emphasised the breadth of construct validity as "an overall evaluative judgment of the degree to which evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions on the basis of test scores….."(p.741). Construct validity is the process of developing a psychological measure and it is embedded in this research from the earliest stages.

Now that the need for an assessment measure for crisis has been identified (section 1.1) the next step is to start developing the measure. Cronbach & Meehl (1955) suggested that there should be three steps toward the development of a measure including a description of the theoretical model, for example the constructs to be described, and how they relate to each other (justification), the development of the measure itself which measures the construct of interest (crisis) and finally the empirical testing of the expected outcomes against the observed outcomes (is the measure actually doing what we want it to do e.g. is it measuring what we want it to measure?). These steps will be followed for the development of the crisis measure in this research. Messick (1995) indicated the importance of understanding the generalisability and the consequences of score interpretation for a newly developed measure which emphasises the importance of developing a clinically credible and practical measurement tool for crisis. It is recognised that as a result of sample variation and measurement error, measurement development is a continuous process of development and evaluation that will never meet a conclusion but move further toward and further away from the chosen goals depending on a number of factors. One of the main reasons this goal will never be fully realised is simply due to the fact that the context and the population will continuously evolve and change over time and as a consequence so too will the expectations and requirements of the measure. In summary, the development of a crisis measure will be a continuing process of development and evaluation with the valued goal direction being to achieve a measure that is as accurate and reliable as possible (Campbell, 1990) for the crisis population of interest at that time. This research will form the first firm step forwards in this journey for crisis measurement.

Simms & Watson (2007) detailed a construct validity approach to measurement design and development (Figure 1.4) based on Loevinger's (1957) outline of construct validity. The approach mirrors the steps taken in measurement development and aims to support construct validity starting at the point when the idea is conceived, through identifying the substance of the construct with the development of the item pool, identifying the internal structure of the item pool, selection of items and finally statistical testing of the reliability and validity of the developed scale. This approach is one of the most comprehensive models for measurement development available and will therefore provide the template upon which this research is based and from which the assessment tool for crisis has been developed. The first phase is centred on substantive validity (Loevinger, 1957), also known as content validity, and focuses on the initial conceptualisation of the construct under investigation and the development of the initial item pool. The item pool selected is accepted to be a sample of all the possible items that could be included in the scale and therefore it is absolutely crucial that it is a representative one. This makes it crucial that the researcher selects a suitably comprehensive approach to item pool development to ensure that it is representative of the underlying construct. For this reason, it is preferred for the item pool to be developed using overly inclusive techniques rather than techniques that may miss, ignore or disregard potentially important items.

Figure 1.4 Simms & Watson (2007) Construct Validity Model

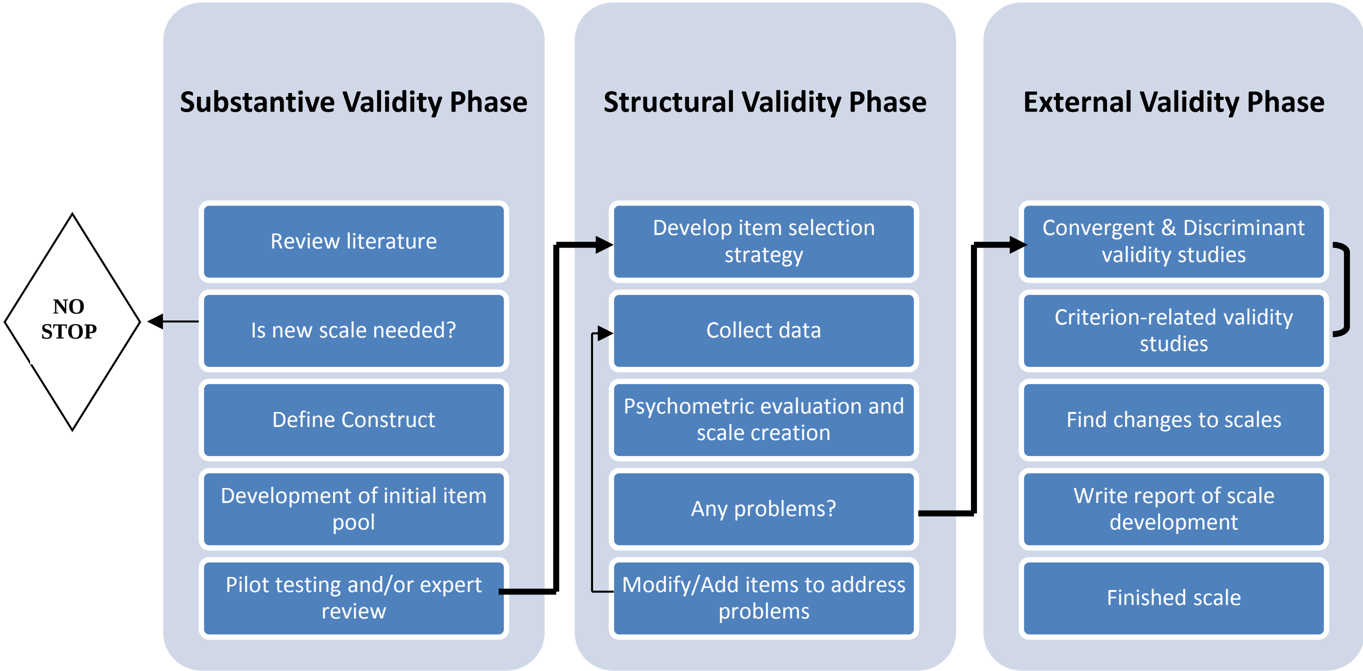


Figure 1.4 Simms & Watson (2007, p.243) Construct Validity Model outlining the three main phases of measurement development to achieve good construct validity.

1.1.6.3 Developing the Initial Item Pool

Single item measures for latent psychological constructs as opposed to measures based on item pools can result in considerable measurement error and therefore psychological constructs require a number of items to help average out the measurement error and reduce the impact of this error on the outcome score (Nunnally & Bernstein, 1994; Spector, 1992). In addition, single items are not able to discriminate and give the precision that multi-item scales offer and often lack the scope to represent the complexity of a psychological attribute such as acute mental health crisis (McIver & Carmines, 1981). With this understanding in parallel to needing to know the exact substance and structure, it is clear that the development of an item pool is a vital first step in psychometric measurement development and therefore for this research also.

"No existing data-analytic technique can remedy serious deficiencies in an item pool" (Clark & Watson, 1995, p.311). This suggests that particular care must be paid at the very early stages of measurement development. The item pool developed must be comprehensive enough to represent the construct under investigation but not too long that the data necessary for validation can never be collected. Over-inclusiveness is expected at this stage of measurement development. However, the balance to ensure that the scale will both be completed and not cause unnecessary fatigue for the participant must also be considered. The main aim of developing a comprehensive item pool is to work towards understanding both the conceptual and empirical boundaries of the construct being investigated. The statistical techniques used later to analyse the items will support decisions regarding items to be deleted or collapsed from the item pool. However, these techniques will not be able to identify items that have been missed out altogether at the item pool formation stage.

The tendency to lean toward over-inclusiveness at this phase is to ensure content validity. Haynes, Richard & Kubany (1995) defined content validity as "…the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose." (p.238). Messick (1989) similarly stressed the importance of over-inclusiveness when attempting to describe a construct or theme with an emphasis on content saturation. The final sample of all the possible trait items will only be as good as the initial item pool for comprehensively covering the content of the trait or latent construct.

It is important to ensure that the item pool is representing the construct in question i.e. that it describes all of the domains/components that make up the underlying construct of crisis. Representation also includes the representation of all levels of the construct of crisis in terms of representing the spectrum of crisis running from low to high. Construct validity at the early stages of item pool development may be obtained through expert review or embedded in the method of developing the item pool e.g. the sample from whom the item pool is developed may represent the expert review.

1.1.6.4 Item Format

The basic principles for writing items and the format for both the item itself and the scale that represents it has been outlined by a number of authors previously (e.g. Clark & Watson, 1995; Comrey, 1988). The two main considerations are item clarity and response format (Simms & Watson, 2007). In terms of item writing, context-neutrality, avoiding bias and item orientation are particular aspects for consideration.

The two response formats most commonly used are dichotomous and polytomous formats. Dichotomous response formats offer just 2 options e.g. yes or no/right or wrong. Polytomous response formats can offer 3 or more response options. The strengths and weaknesses of these two formats have been well documented and therefore will not be outlined here (Clark & Watson, 1995; Simms & Watson, 2007).

1.1.6.5 Pilot Testing

Once the initial item pool and response format has been developed a small pilot should be completed to uncover any obvious design flaws. For example, item wording or scale format can be quickly rectified from feedback received from a small pilot group. If there are no problems identified that would interfere with the data obtained in the main test sample, then this data can always be included in the final item pool and therefore would not be wasted. Therefore, it was an important step in this research to pilot the measure in order to iron out any initial flaws.

1.1.6.6 Method of Item Choice

The structure of the scale will be valid (structural validity) when the items included in the scale represent the breadth and magnitude of the construct under investigation. To appreciate the extent to which the measure developed through this research is able to achieve this goal, consideration will be given to the item development and selection phase. This focuses on how items are chosen and evaluated to ensure that they represent the crisis construct of interest.

The rational-theoretical approach to item selection is when "…the scale developer simply writes items that appear consistent with his or her particular understanding of the target construct, assuming of course, that this understanding is completely correct." (Simms & Watson, 2007 p. 247). If we are assuming that the construct being measured is latent and most likely has not been measured before, it is easy to see the limitations of this approach. At best it is unrealistic to assume that a person or several people's perspectives of a construct can represent the sample or population perception of the construct. Linking back with the theory regarding content saturation (Messick, 1989) can it be said with confidence that content saturation has been achieved using this approach?

Empirical criterion-keyed item selection approach is based on the ability of items to discriminate between a 'normal group' and a 'clinical' or 'criterion' group. This would require identification of both the 'normal' and 'criterion' group prior to item development. The emphasis is not on the development of the items but on the response to items in the pilot phase (Meehl, 1945). This approach has been found to have low internal coherence, high correlations, poor discrimination (Simms, Casillas, Clark, Watson & Doebbeling, 2005) and moves away from the continuum model, back towards a model of diagnosis.

Internal consistency approaches may include approaches embedded in CTT, factor analysis and Rasch analysis which are described in further detail in section 1.2 of this chapter. The overall goal of this approach is to select items that are relatively homogenous and provide good discriminant validity. When utilised, this approach often uses factor/component identification techniques followed by psychometric approaches embedded in CTT and/or IRT to refine the factors/components to produce the final scale item list. This approach will be used for the purposes of developing the crisis measure in this research.

1.1.6.7 Full Measurement Pilot and Psychometric Analysis of Scale Items

The methods used in the internal consistency approach are most common and most appropriate for use in this research as outlined above. To complete this analysis a full pilot of the measure is completed on a substantial sample of the population (Chapter 2) to obtain the data necessary.

1.1.7 The 'Matter' of Crisis

This section has described in detail the importance of sound construct validity and possible approaches and methods for achieving this in the development of a measure for the assessment of crisis. Based on the theoretical underpinnings described in this section, this research takes a comprehensive and exhaustive approach toward the development of an item pool believed to be representative of the latent crisis construct. The initial step in the development of the item pool will aim to be over-inclusive in line with the recommendations outlined in section 1.1.6.3. The importance of appropriate item rating-scale development has also been adequately addressed through this research to ensure that the items are provided with an appropriate method for providing representative information on the level of the crisis construct. This will lay the foundation for the development of a crisis measure with good construct validity.

1.1.8 Common Terminology

Throughout the research there will be the repeated use of a number of common measurement terms that may have been previously discussed but shall be defined here for clarity and reference:

1) Acute Mental Health Crisis - Crisis
2) Categories – these are the points along an item rating scale that indicate the different levels of the item. Categories normally run along a likert style scale of categories that are ranked for the purposes of analysis.

3) Classical Test Theory (CTT) – this refers to the more traditional approaches to measurement development as described in section 1.2.4. CTT approaches fundamentally differ from the more modern approach of Rasch measurement due to its sample dependence.

4) Crisis Resolution and Home Treatment - CRHT

5) Global Overall Measure – this refers to the crisis measure developed through this research as a whole including the information contributed by all the items in the measure as identified in Chapter 3(Figure 3.3).

6) Item Level – this refers to the individual items on the global overall measure.

7) Item rating scale/rating scale – this is the individual rating scale that is developed to accompany each and every item on the measure.

8) Measure, Assessment and Tool – the terms measure, assessment and tool will be used interchangeably to refer to the measure being developed through this research.

9) Principal Component Analysis - PCA

10) Subscale Level – this refers to sub-groups of items that provide information on a number of sub-scales or concepts that contribute to the understanding at the overall global measure level.

11) Rasch Analysis – Rasch analysis is a modern measurement technique utilised for this research to extend and enhance the outcomes of the PCA analysis.

1.2 Exploration of Approaches to Developing and Refining Measures At the Measure, Subscale and Item Levels

The main aim of developing a measure for a latent construct such as crisis is to obtain an accurate indication of an unobservable variable. In mental health, measurement outcomes constitute fundamental information upon which clinical support and research findings are based. Collectively the evidence will support improvements in practice and therefore in patient care. With the current lack of standardised assessment tools to support the assessment of crisis (sections 1.0.9), it is difficult to develop our understanding of the effectiveness of current crisis interventions or to trial new ones to expand the crisis evidence base.

Developing a measure that relates to the underlying latent construct of crisis is a delicate balance between theoretical, clinical and statistical requirements. The resulting measure aims to support, if not improve, the ability to accurately assess and indicate an individual's strengths, weakness and needs in order to deliver appropriate support to help that person resolve their crisis in a helpful way. Measurement theories are techniques that help determine how accurate and therefore successful a measure is for indicating the level on an underlying variable. "Measurement Theory is a theory of how the numbers generated by rating scales relate to measurements of the constructs they seek to estimate." (Hobart, Cano, Zajicek & Thompson, 2007, p.1098). However, the very fact that the variables being measured cannot be directly observed results in measurement error, a difference between the observed information and the true level of the construct of interest. Consequently, one of the tasks of measurement theory in this research is to provide an indication of how accurate the crisis measure developed is in terms of how much of a difference there is between the observed information from the crisis measure and the true level of the crisis construct.

Currently there are two main approaches available, CTT and Item Response Theory (IRT). Combining techniques from these two approaches supports the researcher to obtain information at the level of the overall global measure, subscale and item level. CTT and IRT are both methods that aim to understand the properties of a measure with the aim of ensuring that the measure is doing what is expected of it, in this case, supporting the assessment of crisis. Measurement design techniques have progressed significantly over the last century. CTT encompasses techniques that have been in use by measurement developers for well over 80 years and still continues to make significant contributions to the academic literature. IRT (and in particular the Rasch techniques used in this research) is a modern measurement approach that has recently come to the forefront of measurement design, supplementing and supporting those techniques embedded in CTT.

The main difference between the two theoretical approaches is the level at which they focus in on the measure. There are three main levels of focus (Figure 1.5):

1. The item and rating scale level: this is the level of the individual items in the item pool. This includes the rating scales assigned to items that allow assessors to rate them. This is the focus level of IRT models such as the Rasch model.

2. The subscale level: this is the level at which items are grouped together by their similarities to provide information on a sub-theme of the measure and is a focal level for both IRT and CTT.

3. The global overall measurement level: this is the level where the information provided by the individual items is understood collectively as a whole. This is the main focal level for CTT.

Figure 1.5- Measurement Levels

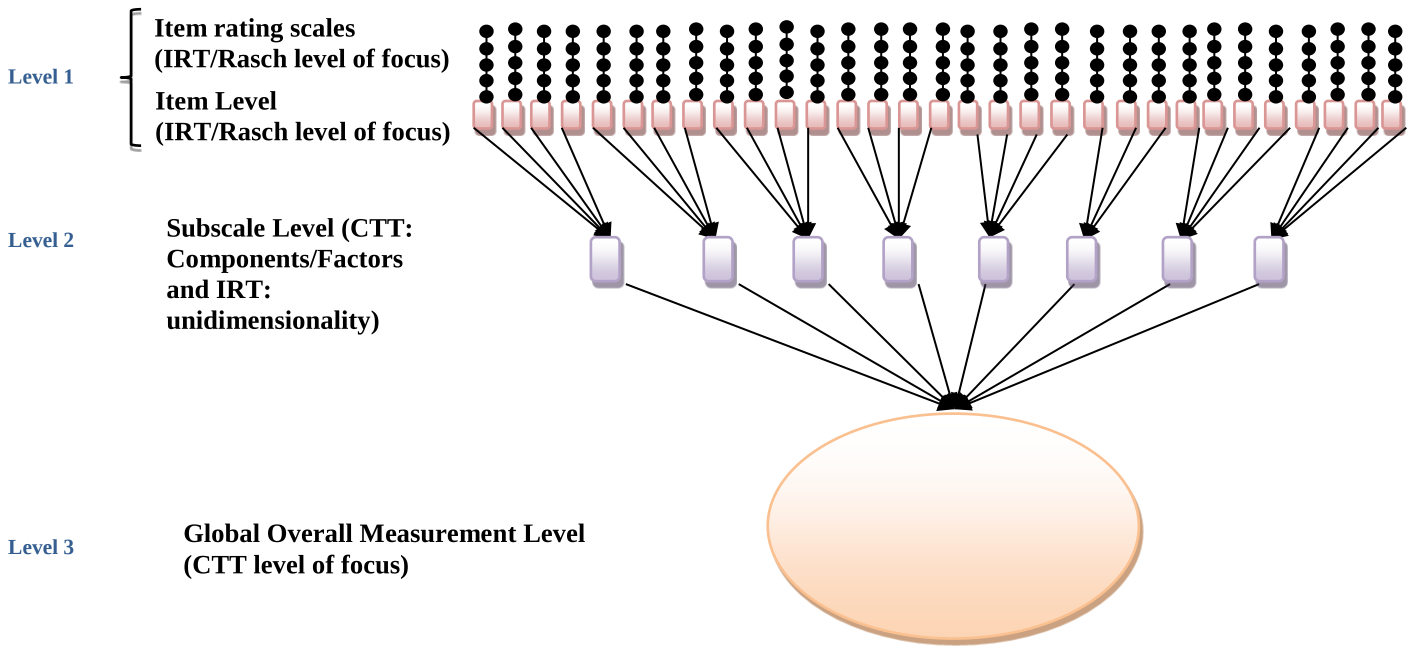


Figure 1.5. Figure outlining the levels of measurement including the item, subscale and global overall measurement levels which will be referred to in this research.

Item Response Theory/Rasch analysis focuses in at an item level to understand individual item contributions and interactions within the measure whereas CTT focuses in at the overall outcome score level to understand what this represents in relation to the construct under investigation. “The former (IRT) focuses on characteristics of test items and how they combine to make tests. The latter (CTT) assumes that a test has been constructed and focuses on the characteristics of test scores.” (Reckase, 2009, pp.62-63). This section aims to understand how both measurement theories will contribute to the development and understanding of a valid and reliable assessment tool for crisis taking into consideration both the advantages and disadvantages of both measurement models.

1.2.1 Measurement Analysis at the Item Level

Modern measurement methods have increasingly been used to supplement traditional approaches to validity and reliability (Tennant & Conaghan, 2007). These approaches are found embedded in Item Response Theory (e.g. DeMars, 2010) with a focus on a related model called the Rasch model (Rasch, 1960) which is used in this research. These techniques are now well established and being increasingly applied in research as demonstrated in the number of published articles that report the outcomes of Rasch analysis (for example; Cano, Barrett, Zajicek, & Hobart, 2011; Fletcher, Kupshik, Uprichard, Shah & Nash, 2008; Pallant & Tennant, 2007).

Item Response Theory models show the relationship between the person’s ability or trait on the underlying construct of interest and the item response (DeMars, 2010). As outlined in Chapter 2 (section 2.6) for the purposes of this research the focus will be on an individual’s *cause for concern* on the underlying construct. Originally the Rasch model was developed in the context of education and so the literature focuses on a common language relating to a person’s *ability* in terms of intellectual ability. The context of this research is different and therefore the term ability will be used for the initial descriptions of this model and then related to the *cause for concern* trait being measured when the empirical research on crisis here is referred to.

The Rasch model is an IRT related model that tests the outcomes of a measure against the mathematical Rasch measurement model (Rasch, 1960), a one parameter logistic model. An IRT model described as a one parameter logistic model is a model that has only one item parameter, namely item difficulty. The two parameter logistic model includes both item difficulty and item discrimination parameters. The three parameter logistic model has an additional asymptote or 'guessing' parameter. The Rasch model (a one parameter logistic model) has a number of advantages for measurement development due to its simplicity, for example unlike the two parameter logistic model and three parameter logistic models, where people who receive the same score are automatically indicated at the same level of ability. The Rasch model only requires the 'number correct' score to give an estimation of ability. The 'number correct' is simply the number of items which were answered correctly in terms of educational assessment. In relation to psychometric measures, these would be the items that were rated to indicate a level on the symptom or presentation. The two and three parameter models are not as simple as this and would over complicate the measurement development process for a newly evolving measure. The differences between the models are complex and will not be discussed further in detail here as they have been well described in the literature elsewhere (for example, DeMars, 2010).

By applying the data collected from the pilot for the crisis measure to the Rasch model, it was possible to understand how the items in the measure were behaving. For example, it was possible to examine item responses to understand whether or not the measure was acting as an interval level scale (an assumption for parametric analysis) and to identify if there was one or several underlying dimensions/constructs (linking in with the underlying assumption of unidimensionality).

Rasch models have been developed for dichotomous measures (Rasch, 1960) with a choice of 2 responses and polytomous measures (Andrich, 1978) that offer 3 or more response categories. Item scores on a measure are added together under the assumption that their summed score accurately represents the level on the underlying construct. To test the validity of this assumption, the item responses are tested against the Rasch model to see if the pattern of responses reflects the model's expectations. This expectation is based on a flexible version of the Guttman scaling model which is a deterministic model (as outlined in Bond & Fox, 2007). The Guttman model suggests that item responses should follow a strict pattern and therefore if a person can score on a more difficult item, they should score on all the easier items before it. The Rasch model is a more flexible version of this whereby if a person scores on a more difficult item they have a high probability of scoring on easier items but this is only a probability and not an expectation. In attempting to measure a construct as complex as crisis a more flexible scoring approach, based on probability rather than certainty, would be more preferable and more realistic.

1.2.1.1 Item and Person Estimates

The probability approach described is a "*logistic function of the relative distance between the item location and the respondent location on a linear scale*" (Tennant & Conaghan, 2007). In basic terms, a logistic function is an S-shaped curve which was developed by Verhulst in 1845 in relation to population growth (Weisstein, 2011) with the concept that initial growth is approximately exponential and as saturation begins, the growth slows (Figure 1.6).

Figure 1.6 S-Shaped Logistic Function

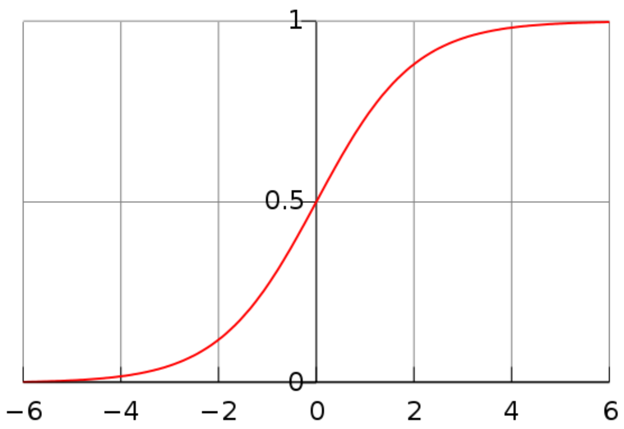


Figure 1.6 outlining a S-shaped growth logistic function curve upon which Rasch modelling is based. This shows slow initial growth, followed by exponential growth before finally the growth slows and levels off.

The item and person locations in terms of Rasch analysis are their locations on a linear scale that represents the underlying construct in terms of the person's ability and the item's difficulty. The location of an item on the scale is the point that corresponds with a probability of 0.5 for a 'correct' response (a 50:50 chance). Therefore, where a person's ability is >0.5 they have a greater probability of correctly answering that item or receiving a rating. Placing this in the context of more psychological measures such as the Beck Depression Inventory (Beck, Ward, Mendelson, Mock & Erbaugh, 1961) a person's location would represent the likelihood of them scoring highly for depression. Instead of calling this 'ability' it would be their 'level of depression' and in the context of crisis assessment it would be their level of crisis or *cause for concern* in the specific context of this research.

"In other words, the probability that a person will affirm an item is a logistic function of the difference between the person's level of, for example, pain and the level of pain expressed by the item, and only a function of that difference." (Tennant & Conaghan, 2007, pp.1358-1359). The person and item estimates are subjected to a log transformation and are often displayed against a logit scale (log odds unit scale) (Bond & Fox, 2007). This logit scale is an interval level scale and therefore the units on the scale have a consistent meaning (interval level scaling) allowing individual people or items to be compared. This is sometimes referred to as the person or item location (on the logit scale). The item and person information is separated out and can be studied independently but at the same time is placed on the same metric which allows comparison. By separating out the item and person information, the item information is more independent of the sample and as a consequence can be more easily generalised.

1.2.1.2 Rasch and the Assumptions of Interval Level Scaling for Parametric Analysis

The purpose of developing a measure for crisis is to provide outcomes that are meaningful and can be compared over time or against normative data. In addition, the measure would provide a valid method of collecting data for the purposes of research which ideally would be analysed using parametric analyses.

As described earlier, in the physical sciences it is possible to make direct measurement. However, this is not possible in mental health and the challenge is to indirectly measure latent variables in a meaningful way that gives a true representation of the 'amount' present. Rating scales are a method for achieving this and a method that is widely adopted in both clinical practice and research. In mental health multi-item scales are used as a useful method for breaking complex variables down into their component parts which cover the construct of interest (Hobart, Cano, Zajicek & Thompson, 2007). The crisis assessment construct of interest for this research will be broken down into a number of individual items that are anticipated to collectively contribute information to an overall understanding of the construct. In addition to the items identified to represent the construct, rating scales are developed to indicate a level on each of the items in polytomous measure, made up of a number of ordered categories. There are two assumptions – 1. The items comprehensively and accurately represent one underlying latent variable/construct (unidimensionality) and 2. The rating scale represents accurate interval level measurement rather than simply ranked categories. These are the assumptions of parametric analysis. Until recently, the majority of rating scales in psychometric measurement were designed using ordinal level scales. Ordinal level scales are a list of categories that are ranked in an order. For example, the Beck Depression Scale (Beck, Ward, Mendelson, Mock & Erbaugh, 1961) has four categories:

(0) I do not feel sad.

(1) I feel sad.

(2) I am sad all the time and I can't snap out of it.

(3) I am so sad or unhappy that I can't stand it.

The numbers assigned simply rank the categories on an ordinal scale but do not indicate that the distances between the categories are equal (as is the case with interval level scaling). For example, the psychological leap from category 1-0 may be far greater than the leap from category 2-1. However, these measures are often treated as if they possess interval level scaling for the purpose of research. Therefore, measurement tools that use ordinal scales are difficult to interpret in a meaningful way.

Interval level scales are scales where the distances between categories are constant across the metric – "The analysis and interpretation of differences in scores and changes during time are most meaningful when the unit of measurement is constant and the numerical meaning of the numbers is maintained when they are subjected to statistical analysis."(Hobart, Cano, Zajicek,& Thompson, 2007). Figure 1.7 shows a comparison between ordinal level and interval level data. In terms of parametric statistics, it is argued that only interval level data is appropriate for this form of statistical analysis and ordinal data should only be analysed using non-parametric alternatives (Stevens, 1946). Therefore, the aim of modern measurement developers is to provide interval level scales to provide meaningful outcomes. This can be achieved with the support of modern measurement methods embedded in the Rasch model "..the aim is to provide social scientists with the means to produce genuine interval measures and to monitor the adherence of those scales to scientific measurement principles, so that Rasch estimates of ability/attitude/difficulty become the data for statistical analysis." (Bond & Fox, 2007. p.5). For the aims of this research to be achieved requires the development of an accurate measure which can only be assumed when interval level scaling is present.

Figure 1.7 Ordinal Level Scaling Compared to Interval Level Scaling

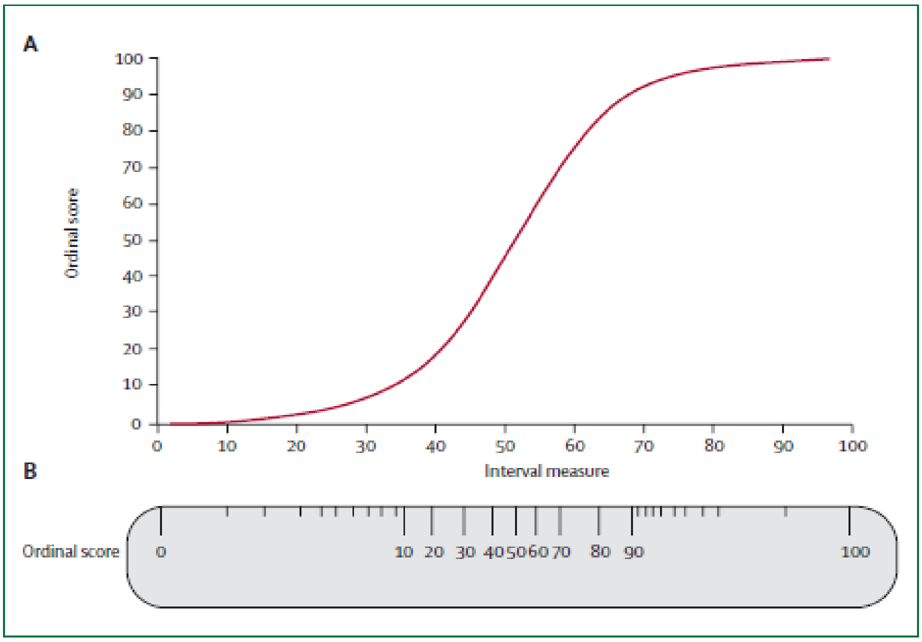


Figure 1.7 showing the comparison between ordinal and interval level data resulting in an S-Shaped curve. Figure taken from Hobart, Cano, Zajicek & Thompson (2007, p.1099).

1.2.1.3 Internal Consistency and Unidimensionality

The internal construct validity of a measure is accessed through the assessment of the measure's unidimensionality. It is assumed that the items in a measure relate to one underlying latent construct of interest and therefore the summed score of the items represents a level on this construct. If the items on a measure relate to more than one underlying construct then any summed score will be contaminated by influences from the other constructs they represent as well as the one of interest. Therefore, the validity of a measure is partly assessed by whether it is unidimensional and this will be a fundamental consideration in developing a crisis measure through this research.

Internal consistency and unidimensionality are often confused to be the same thing. Internal consistency is necessary if unidimensionality is to be achieved (Clark & Watson, 1995; Schmitt, 1996) but they are not the same. Internal consistency is the extent of inter-relationships between items whereas unidimensionality (also known as homogeneity) looks at the extent to which the items all relate to the same underlying construct.

The Rasch model holds the assumption of unidimensionality. One of the most common methods for assessing unidimensionality is a Principal Component Analysis of the residuals as proposed by Smith (2002). In basic terms, once the Rasch model factor has been taken into consideration there should be no further associations between items apart from by chance. Where patterns or relationships between items are not shown between the residuals, a unidimensional measure is demonstrated. Smith's (2002) model examines the relationship between the items and the first residual factor which helps to identify two subsets of items with one positively and one negatively correlated. The person estimates are calculated for each group and if unidimensional it is expected that they will produce similar outcomes. If the items are not unidimensional it would be expected that significant differences between the two sub groups of person estimates would be observed. An independent t-test is used to test for significant differences between the two subsets of person estimates for each person and the percentage of significant differences is calculated and expected to be <5% if the measure is unidimensional.

The Rasch model provides a useful approach to assessing unidimensionality that will be adopted to test the crisis measure developed. If the measure is shown to be unidimensional then the summed score of the measure will be able to provide a valid and meaningful indication of the level on the construct.

1.2.1.4 Response Dependency

The Rasch model also has the underlying assumption of local independence which means that items should not be related to each other, apart from due to the underlying construct. Once the Rasch factor has been extracted (i.e. the main scale) there should be no left over patterns demonstrated by the residuals. The popular example of item dependency is in a measure of mobility that contains the items asking ‘Can you walk 1 mile’ followed by the question ‘Can you walk 50 meters’. It is obvious that if you have a positive response to the first question, you will have a positive response to the second question and therefore these items demonstrate local dependency and suggest scope for item extraction. The benefit of assessing for local dependency is that items shown to depend on each other after the Rasch factor has been removed can be assessed and removed where appropriate which will reduced the item pool and avoid information repetition. This will be a powerful tool for developing a crisis measure through this research helping to reduce any redundant items. As the research will initially focus on being over-inclusive, the aim will be on reducing the item pool down to a more manageable size whilst retaining the relevant information of interest for measuring the crisis assessment construct.

1.2.1.5 Reliability at the Item Level using Rasch

The reliability of the developed measure from this research will be important for integrity if it is to be used in the community by CRHTs. The Rasch model provides a measure of reliability based on an index of how accurately the scores achieved separate or discriminate among examinees. This appraisal of reliability is called the Person Separation Index (Wright and Masters, 1982) and is based on the CTT principle of the ratio of true score variance to observed variance (DeMars, 2010). This is equivalent to Cronbach’s alpha (1951) but uses the logit value rather than the raw score as used in CTT (as outlined in section 1.2.3.2). In line with Cronbach’s alpha it is expected that the Person Separation Index should be $\geq$0.7. It is important that the outcomes of the research’s measure are able to provide reliable and therefore accurate information about a person’s level of crisis. The Person

Separation Index will offer an indication of how reliable and accurate the crisis measure is and therefore will be a valuable tool in this research.

1.2.1.6 Choosing a Rasch Model

The Rasch model has a significant role to play in this research. There are three main Rasch models to choose from and it is important to make the right decision before carrying out any Rasch analysis to ensure that the results obtained are representative and meaningful. The Dichotomous Rasch model is used when there are only 2 response options and as indicated in Chapter 2 will not be used for this research. When there are 3 or more item-scale response categories there are two potential models to choose from, the Andrich Rating Scale Model (Andrich, 1978) or the Masters Partial Credit Model (Masters, 1982). Both approaches use the Rasch model but the mathematics differ slightly depending on the model used. The Andrich Rating Scale Model assumes that the distances between the thresholds of the item rating scales are equal i.e. the steps between the categories of the item rating-scale are expected to be equidistant. The Masters Partial Credit Model assumes that the distances between the thresholds and therefore the categories are not equidistant. The name 'partial-credit' has its origins in multiple choice questions where some 'incorrect' answers still indicate some knowledge and therefore give partial credit toward the correct answer. The RUMM2030 software package (2010) for used in this research for Rasch analysis offers a likelihood ratio statistic that can guide toward which model to choose.

When choosing to adopt the Partial Credit Model it is important to examine the item rating-scale category structure. Where categories are shown to have ordered thresholds, therefore indicating clear steps along the underlying construct, the categories are healthy and provide useful information. When the category is redundant it is not shown to be the most probable option along the continuum at any point, not able to contribute any further information than that already provided by the other categories. When a redundant rating scale category is identified, it is possible to collapse these categories down into other healthier categories. This process is outlined in Chapter 4 in the context of the development of this crisis measure.

1.2.1.7 Test of Fit to the Model

By applying the data collected through this research to the Rasch model it is possible to obtain an estimate of how well the data, separately for persons and items, fit the Rasch model. This statistic is simply called the fit statistic. This provides a useful approach for reducing redundant items and removing persons that are acting as outliers and would therefore skew the outcomes of analysis. The fit statistics used in the Rasch model are chi-square based whereby the focus is on the difference between the observed response and the Rasch expected model outcome across groups of different ability. If the outcome of the chi-square analysis is <0.05 then the item is deemed to misfit the model's expectations. The RUMM2030 (2010) software used in this research produces a residual statistic as well as the chi-square statistic. The residual statistic is the standardised sum of all the differences between the observed outcome and the Rasch expected outcome summed across the whole sample. The fit statistics for both persons and items will support the development of a valid and reliable crisis measure by examining the pattern of information observed against the Rasch expected model.

1.2.1.8 Reporting Expectations in Rasch Analysis

Tennant & Conaghan (2007) suggested reporting criteria that cover 7 fundamental aspects of Rasch analysis. This will provide a useful checklist against which this research can be compared and used for reporting the outcomes of the analysis :

1. The model chosen
2. Where polytomous, the appropriate ordering of the categories and any necessary rescoring
3. Fit of items and persons to the model and justification for the fit levels chosen, strategy for improving fit (e.g., item deletion) and resulting fit statistics.
4. Local independence of items including response dependency and unidimensionality
5. Differential Item Functioning
6. Targeting of the scale
7. Person Separation reliability

Checklist point number 5 is for Differential Item Functioning which will not be assessed through this research but is discussed further in the Discussion section (Chapter 9).

1.2.1.9 Item Difficulty

In Classical Test Theory (CTT), item difficulty is the proportion of people in a sample who 'correctly' answer an item compared to those who answer it incorrectly. In general, item difficulty levels that support the item to differentiate between the populations of interest are statistically most useful e.g. items that are closest to 0.5 therefore differentiating the population 50:50 are most helpful. Items that have a difficulty of either 1.00 or 0.00 are redundant items as they indicate that either 100% of the population answered correctly/endorsed the item or 100% of the population answered the item incorrectly/failed to endorse the item. In reality, test developers attempt to set scales that have varying difficulty values across items to differentiate between more ability levels but also to ensure that test takers do not become disheartened or give up if they are of lower ability level (Ghiselli, Campbell, & Zedek, 1981). Item difficulty is a useful statistic as it supports the measurement developer to identify items that represent the underlying construct of interest, in this case the construct of crisis, across the entirety of the construct spectrum. Therefore, a person's level on the construct can be indicated by the measure's outcomes. For the purposes of differentiating between individuals across the spectrum of the crisis construct this is a particularly powerful tool and will be a useful analysis in this research.

1.2.2 Subscale Level Analyses

Following the identification of the substance (item pool) of acute mental health crisis, sense is made of the item pool by identifying its structure. This investigates how items relate to each other to identify the components or subscales of the item pool.

1.2.2.1 Exploratory Factor Analysis & Principal Component Analysis

There are a number of techniques available for identifying the structure of the crisis measure's item pool that may help to identify subscales. The data collected from rating the items in the item pool can be analysed using Exploratory Factor Analysis (EFA) or PCA, both techniques associated with CTT. There are three important decisions to be made when completing EFA and PCA analysis (Preacher & MacCallum, 2003). 1) Which model should be used (exploratory factor analysis or PCA), 2) the number of factors/components to retain, and 3) the rotation method. These decisions are important for ensuring that the outcomes of analysis will be meaningful and representative of the underlying crisis constructs. In the context of developing a crisis measure this would be the step towards understanding how the items identified from the item pool can be assessed for their relationships to each other. This may result in the emergence of smaller sub-groups (or subscales) that make up the structure of the overall group of variables in the item pool. This will be a useful step in the development of the crisis measure and in identifying the key areas of crisis assessment. The items identified for each subscale will describe in detail the areas for crisis assessment which will provide useful guidance for crisis teams for completing comprehensive assessment.

1.2.2.2 Choosing a Model

Exploratory Factor Analysis (EFA) and PCA are often used interchangeably as they are mistakenly believed to deliver similar outcomes. However, there are significant differences between these two approaches that need to be considered for the purposes of choosing a model for this research (Preacher and MacCallum, 2003). Firstly, the primary aim of EFA is to identify the underlying latent variables (Comrey, 1988). PCA on the other hand is a method of data reduction. EFA separates its data into those that share common variance (variance accounted for by common factors) and unique variance (variance that is not attributable to common factors). PCA does not try to differentiate common variance, unique variance and error when identifying components. Therefore factors and components are statistically different outcomes. In basic terms PCA is a method of grouping together variables/items by how similar they are to each other and how dissimilar they are to others. By studying the patterns of correlations between variables it can be identified where a number of variables may be measuring aspects of the same underlying dimension or component. As outlined by Field (2009, p.628) the PCA technique has three main uses; 1. to understand the structure of a set of variables, 2. to construct a questionnaire to measure an underlying variable and 3. to reduce a data set to a more manageable size while retaining as much of the original information as possible.

With the aim of over-inclusiveness for the initial item pool development, it will be helpful to utilise PCA techniques to reduce the item pool in the following stages whilst identifying the structure of the crisis measure therefore making it the approach of choice for this research.

1.2.2.3 Number of Factors/Components to Retain

The decision for retaining components is one based on theory, statistical guidance and clinical understanding. There are a number of statistical methods available to support the researcher in making a decision about the components to retain within a scale structure. These decisions are made in the knowledge of the context of the measure and the practical and clinical implications these decisions will have on the scale. This is an important step in the process of developing this crisis measure as it may result in further reductions in the item pool and it is important to ensure that the subscales essential to the understanding of the underlying crisis construct are retained. The possible approaches to component or subscale retention are outlined here:

1. Subjective assessment of the scree plot (Gorsuch, 1983). The scree plot is a graph plotting the eigenvalues in order of magnitude and when assessed can identify the number of components to retain based on the number of components that lie before the last steep drop on the graph. Tzeng (1992) and others before (e.g. Cattell & Vogelmann, 1977) have found this to be a reliable method.

2. Parallel analysis of eigenvalues (Horn, 1965; Humphreys & Ilgen,1969) is based on identifying the components that account for more variance than could be expected by chance. An equation developed by Montanelli and Humphreys (1976) helpfully indicates the value of the leading eigenvalues based on the sample size and the number of items.

3. The Kaiser-Guttman rule for retaining components guided by a lower bound eigenvalue of 1 (Guttman, 1954; Wainer, 1982). Although widely adopted in research the reliability of this method has been questioned and shown in research to both overestimate (Zwick & Velicer, 1982) and underestimate (Humphreys, 1964) what should be retained. In addition, it is often found in the literature that researchers mistakenly assume that all factors/components with

eigenvalues greater than 1 should be retained when really this is simply marking a lower bound.

4. The maximum likelihood parameter estimation which is associated with a number of fit indices including the likelihood-ratio statistic and the Tucker–Lewis index (Tucker & Lewis, 1973). This is a special case of Structural Equation Modelling and therefore will not be described in detail here.

Parallel analysis has been shown to be the most accurate and will therefore be the approach of choice for this research. Although these methods provide a guide for component retention the resulting components must be meaningful both theoretically and psychologically for the purposes of psychometric measurement. Therefore, for the purposes of this research the statistical approach chosen will simply act as a guide to component retention in addition to the theoretical and clinical understanding of the components identified.

1.2.2.4 Rotation Method

Rotation is a method for obtaining a perspective that allows the strongest characteristics of each item to be emphasised (DeVellis, 2006). The two main methods of rotation are oblique rotation and orthogonal rotation. Orthogonal rotation assumes that the resulting components will not be related whereas oblique rotation does not make this assumption. The emphasis is on the researcher to pre-empt whether or not there will be correlations between the components and to choose an appropriate technique, which of course is not possible. Therefore, it is wise to complete an oblique rotation first to assess if there are correlations between the components and where there are to proceed using oblique rotation and where correlations are not demonstrated, to proceed with orthogonal rotation.

1.2.2.5 Retaining Items Based on Loadings

Deciding the cut-off for item loadings onto components is crucial and can be the difference between an item being extracted from the item pool altogether or retained. This will be important for the development of the crisis measure to ensure that a representative and comprehensive item pool is retained. Extracting an item unnecessarily could result in the scale losing crucial information that may potentially impact the overall outcome of the measure and therefore impact clinical decisions made.

Although item loadings guide the researcher toward which items should be extracted, it should be remembered that the loadings are affected by sampling error and therefore may not transfer to other samples of the same population, acting only as ‘guidance’ and not replacing clinical judgement and experience. The loading cut-off will depend on the field of research (Preacher & MacCallum 2003). For pure sciences cut-offs of 0.7+ may be expected but for social sciences such as Psychology cut-offs of 0.3 and 0.4 may be equally as meaningful. Principal Component Analysis has been chosen for the purposes of this research as a variable reduction approach and therefore guidance from the literature regarding the cut offs for PCA loadings in social sciences. This will be described in further detail in Chapter 3 where the outcomes of the PCA analysis for this research are described.

1.2.2.6 Component Names

Although EFA or PCA techniques can guide the researcher toward component retention decisions, there is no statistical technique that can support decisions for naming the resulting components. The aim for the researcher is to obtain consensus from experts in the field whether professional experts or experts by experience to make the final decisions regarding a name that best represents the theme of the subscale. This research uses the expert opinion of the research team and others involved in the research including CRHT staff and patients to decide upon appropriate names for each of the subscales identified through the PCA analysis (Chapter 3).

1.2.3 Analyses at the Global Overall Measurement Level

Classical Test Theory focuses in on the meaning of the total measure score, crucial to the final stages of measurement development to assess how the scale is functioning. It is particularly interested in reliability and validity of the measure (Loevinger, 1957). Validity focuses on the ability of the measure to accurately represent the construct of interest, in this case, crisis. The reliability of the measure focuses on the ability of the measure to accurately indicate a level on that construct that is stable and reliable. Both the validity and reliability of the crisis measure developed through this research will be essential for the integrity of the measure and will provide evidence to support CRHTs in adopting it for assessment in their services.

1.2.3.1 The Underpinning Assumption of Classical Test Theory

Before outlining the statistical approaches offered by CTT, it is helpful to understand some of the underpinning assumptions of this theory. The key assumption held in CTT is that the observed score is made up of both a ‘true score’ and ‘random error’. The true score is the actual level of the construct of interest (crisis) and the error is the difference between this true score and the level indicated on the measure, therefore the true score remains hypothetical but this is the foundation upon which CTT has been built. CTT assumes that the error is completely random and therefore has a mean of 0. This helps to explain why a person’s score may vary on a scale that is meant to measure a fairly stable trait such as IQ. It is expected that a construct as complex and changeable as crisis will produce less stable outcomes. This presents a particular challenge to the development of a stable outcome measure. The extent to which this statistical approach has been adopted to understand the functioning of measures is evidenced in the literature which indicates that the techniques embedded in CTT have been useful for communicating the validity and reliability of measures and offering a useful approach for communicating the outcomes of the crisis measurement tool developed through this research.

1.2.3.2 Validity and Reliability at the Total Score Level Using Approaches from CTT

It is important for this research that the resulting crisis measure is both valid and reliable. “…we need to establish that our measuring instrument is *reliable*, that is, consistent, and measures what we set out to measure, that is, the test is *valid.”* (Domino & Domino, 2006). The principal concepts around validity and reliability will be briefly outlined here with a full assessment and analysis of the validity and reliability of the crisis measure developed through this research in Chapter 8.

Validity

As suggested by Strauss and Smith (2009) the fundamental aspects of construct validity are understood based on the classic papers now published over 50 years ago (e.g. Campbell & Fiske, 1959; Cronbach & Meehl, 1955; Loevinger, 1957). These papers are still just as relevant today in terms of the principles they offer to the understanding of measurement validity and, as described by Bornstein (2011), many validity studies published today still look towards these very first understandings of validity to base their research approach. There have been a number of different 'types' of validity presented in the literature but there appears to be a leaning towards Loevinger's (1957) theory of validity that suggested construct validity to encompass predictive, content and concurrent validity. This will be the primary focus of validity for this research also.

It is understood in the literature (Strauss & Smith, 2009; Loevinger, 1957) that validity is not 'proven' within the parameters of a single experiment but instead it is accumulated over time and over the course of several research projects or experiments. It is now accepted that the validity of a measure is something that continuously evolves and develops over time as it aims to get as close to the valued goal as possible. Therefore, construct validation is developed and not established. Based on this understanding, the aim of this research would simply be to take the first steps in developing a valid measure that will be further tested over time.

The difficulty presented for this research is to develop a valid measure without an existing knowledge base regarding the concept or construct of crisis and without already existing measures against which to check the test validity. As succinctly described by Strauss & Smith the challenge for validity when developing a first of its kind measure is that "The goal of validating measures of psychological constructs necessarily requires criteria that are themselves valid." (p.2). This is similar to the challenges faced by some of the first psychometricians and therefore the literature regarding this predicament published over 50 years ago is particularly relevant here but is also accepted in the current literature as still providing the key underpinnings to theory in this area. Therefore, due to the lack of a comparable measurement criterion, it will be important for this research to state from the very first stages what the measurement tool developed through this research aims to measure (defined later in this section).

There has been great progress in the realm of measurement validity from the early conceptualisations of validity, to criterion-validity, through to construct validity (Loevinger, 1957) and theory around convergent and discriminant validation (Campbell & Fiske, 1959). It was during this period when validity first became a relevant and important issue in the literature that the concept of a "hypothetical construct" was first described by MacCorquodale & Meehl (1948). They described a hypothetical construct as a "cognitive factual reference" (p.107), hypotheses about entities, concepts or processes that cannot be directly observed but are still legitimate in terms of measurement. This definition of a construct will be adopted for the purposes of this research.

More recently there has been a focus on four main areas of validity in psychological measurement research; 1) The indeterminate nature of the validation process (Bartley, 1987), 2) The lack of precise strong theories upon which to base the development of psychological measure (Kane, 2001), 3) The evaluation of psychological process theories that explain how participants respond to measures and experiments (Knight & Silverstein, 2001) and 4) The importance of measuring homogenous constructs (Smith et al, 2003). These areas of focus continue to be at the forefront of measurement development in the Social Sciences and in particular Psychology and although there may not be one particular answer to these challenges, it is the aim of this research to work towards answering them and acknowledge the challenges they present to the measure developed here.

It is recognised that this measure will aim to tap into some of the key concepts identified in the theory (as outlined in Chapter 1) including ideas around coping, resources, support, risk factors and meeting basic needs in order to determine whether or not an individual is able to function safely in the community whilst they receive crisis treatment or whether they require inpatient admission. The definition of measurement focus for this research is outlined in section 1.4.1.

Reliability

There are three main forms of reliability that will be considered in the development of this crisis measure. These will be described using the concrete example from physical measurement of the meter rule:

1. A meter rule will not change in length over time (***temporal/test-retest reliability)*** and therefore the measurement of a constant thing such as the height of a house should provide a stable and consistent outcome and accurately represent when change occurs over time.
2. The meter rule may be divided up into smaller units such as centimetres which should also be stable and consistent. This means that the first centimetre is the same as the second and the third and so on. This is known as ***internal consistency reliability***.

3. Lastly, it can be assumed that when the meter rule is used by two different people to measure the same thing the outcomes will be very similar. This is called ***inter-rater reliability.***

1.2.4 Classical Test Theory and Rasch

Classical Test Theory (CTT) is well established in measurement theory as evidenced in a breadth of measurement development studies. Rasch analysis is relatively new when compared to CTT but it is showing itself to be a valuable contributor to the understanding of validity and reliability of measures. Currently there is a trend to support either one approach or the other. However, it seems more appropriate for the purposes of this research to look to the strengths of both these models to enhance the analysis. However, it is helpful to think about the strengths and weakness of both models to hold in mind throughout this research.

1.2.4.1 Advantages and Limitations of Classical Test Theory (CTT)

One of the advantages of CTT is that it is one of the most popular approaches to understanding the quality of measures and therefore is one of the most well known in both clinical practice and research. This is an advantage for communicating the outcomes of measurement design. It is also evidenced in the literature that CTT often yields very similar results to some of the more modern measurement models (DeVellis, 2006). The main concern of CTT is that two of the fundamental components (true score and the error) cannot be determined and therefore the accuracy of the outcomes can never be verified. Hobart, Cano, Zajicek & Thompson summarised 4 main challenges in CTT. Firstly, due to the difficulties in determining meaningful true score and error, the measurement theory is weak and therefore results in weak inferences. Secondly, for the same reason, the theory cannot be challenged and therefore is often easily satisfied. The third challenge is that only the raw scores can be analysed because the parameters cannot be accurately estimated and therefore error is involved in any outcomes observed. Finally, the CTT equation for calculating confidence intervals results in large values which reduces confidence when interpreting change.

CTT is reliant on relatively large samples if the outcomes are to be representative of the population. Therefore the statistical information generated from the sample can only be applied with confidence to that same sample and questions the generalisability of its outcomes. In addition, it is known with CTT that tests become more reliable the longer the test but this often results in repetition of items and increasing rater fatigue which ultimately leads to bias. The true scores are assumed to be measured on the interval level but as discussed earlier there are no means in CTT to test this, which challenges the validity of the outcomes and an assumption of parametric analysis.

1.2.4.2 Advantages and Limitations of the Rasch Model

The advantage of using the Rasch model is that the focus is at the item level. This means that the person and item statistics are separated out and become independent, which has allowed the Rasch model to overcome one of the main flaws of CTT i.e. the outcomes are more generalisable. Another advantage is that the person ability estimates and the item difficulty estimates are plotted along the same metric allowing for direct comparison. In addition, the Rasch model supports the development of much shorter and more concise measures by matching items to person ability. This is demonstrated in Computer Adaptive Testing whereby a subset of items are first administered that represent average ability, with further items administered to gauge if that person is above or below this average level. The advantage of this approach is not requiring all items in a measure to be administered to obtain a meaningful level on the construct. This has led to the development of item banks where large numbers of items are held electronically, representing various points along a construct that can be administered in different sequences to gauge a person's ability level. With a large bank of items, each test could potentially differ to the next reducing practice effects and the potential for cheating between candidates. Developing the crisis measure using approaches from Rasch analysis means that the foundations necessary to enable this type of more complex development could be possible in the future.

One of the limitations of the Rasch model is that it has suffered from its own celebrity and is sometimes misused as a cure for flawed measurement design. It is a useful tool for calibrating item statistics and assessing the construct being measure for example but it cannot substitute high quality measurement design. In addition, it is recognised that the Rasch model does not necessarily produce better measures than those produced using techniques from CTT and therefore it should not be used in isolation but in collaboration with techniques from CTT and other measurement models (Xitao, 1998). Another consequence of its fame is the misunderstanding that where a measure shows statistical strengths in relation to the Rasch model automatically gives the measure clinical strengths in practice. The emphasis is on the experience of the researcher to take the statistical strengths of the measure in balance with the theoretical, clinical and practical strengths of the measure for the purposes of decision making. Afterall, there is no point in having a statistically perfect measure if no one will ever use it because it is impractical and difficult to translate in a clinically meaningful way.

1.2.5 Applying Classical Test Theory and Rasch Analysis to the Development of an Assessment Tool for Acute Mental Health Crisis

Attempts to summarise a complex presentation such as acute mental health crisis using quantitative means will not contain the entire richness and the depth of that experience. The rationale for taking steps to develop a measure for crisis assessment is to present the complex and difficult concepts in an objective and standardised manner to assist in treatment decisions. The complexity, breadth and intensity of the experience suffered by people experiencing crisis is a significant obstacle to comprehensive crisis assessment. Currently, the same person assessed by two clinicians may result in very different assessment outcomes with different elements of the construct of crisis being given attention and yet completely omitted in the other, simply due to the lack of standardised assessment guidelines (Chapter 2, section 2.5.2). Decisions based on these outcomes should be used in the construction of the care plan for the patient. How can the two be compared when different criteria are used? How can the level of crisis be measured when relevant information may have been omitted and irrelevant information included? The development of a valid and reliable crisis tool will provide a more objective and standardised approach to assessment which will guide clinicians towards more appropriate and relevant intervention.

It has been shown that the Rasch model and CTT approaches both have their strengths but also their weaknesses. However, when used in collaboration they provide a powerful approach for comprehensively assessing the integrity and quality of a measure as well as offering approaches to make further refinements and improvements.

The focus of this section of the introduction has been on the use of statistical models to develop and understand the crisis measure developed through this research. However, it is important to stress that no statistical model will be able to make up for the poor identification and design of items and rating scale in the earliest stages of measurement development. Without construct validity, statistical analysis becomes a redundant exercise.

1.3 The Development of a New Crisis Measure – Rationale for Research

What is evident from crisis theory is the complexity and temporary nature of crisis that can be successfully resolved with appropriate support delivered in a timely fashion. In some cases post-crisis functioning is marked by significant improvement from the pre-crisis functioning suggesting that the experience of crisis can in fact improve post-crisis quality of life. The aim of the research in this thesis is to improve assessment and the success of treatment in crisis cases as a result.

To deliver appropriate intervention requires appropriate methods for determining the nature of the crisis. This would include the ability to identify areas of strength and the areas of weakness in order that treatment may be appropriately targeted. Appropriate methods of assessment have been discussed but not developed in a number of crisis models (Aguilera, 1998; McGlynn & Flowers, 2006; Roberts & Lewis, 2001). If appropriate support is to be delivered it is necessary to acquire an accurate and comprehensive assessment in order to design appropriate clinical intervention. Previous attempts to develop a crisis measure have not used comprehensive or systematic approaches to uncover the complexities of the crisis construct or used rigorous statistical analyses to demonstrate the validity (including unidimensionality) and reliability of the scale. Therefore, there continues to be a significant need for the development of a standardised, valid and reliable crisis measure which will be answered in this research.

1.4 Research Aims, Definition of Focus and Overview of the Research Design

The literature review shows that there is a continuing need for the development of a standardised, valid and reliable measure for crisis. This has been hampered by the previous failure to yield evidence that the constructs of crisis assessment have been adequately identified in order to lay the foundations of measurement development. This was, until recently, when the work of Tobitt and Kamboj (2011, section 1.0.2.4) took the first significant steps to rectify this situation through the qualitative investigation of crisis and crisis work. However, they acknowledge that the patient perspective was still missing and would only be complete once this had been included.

This research takes the next steps to develop a measure using a measurement development design approach with a focus on two main aims for the development of a psychometric measure for crisis:

1. To develop a comprehensive and exhaustive investigation into the concepts/constructs of crisis assessment.
2. To develop a valid and reliable measure to support mental health professionals to assess and treat individuals referred to the CRHT.

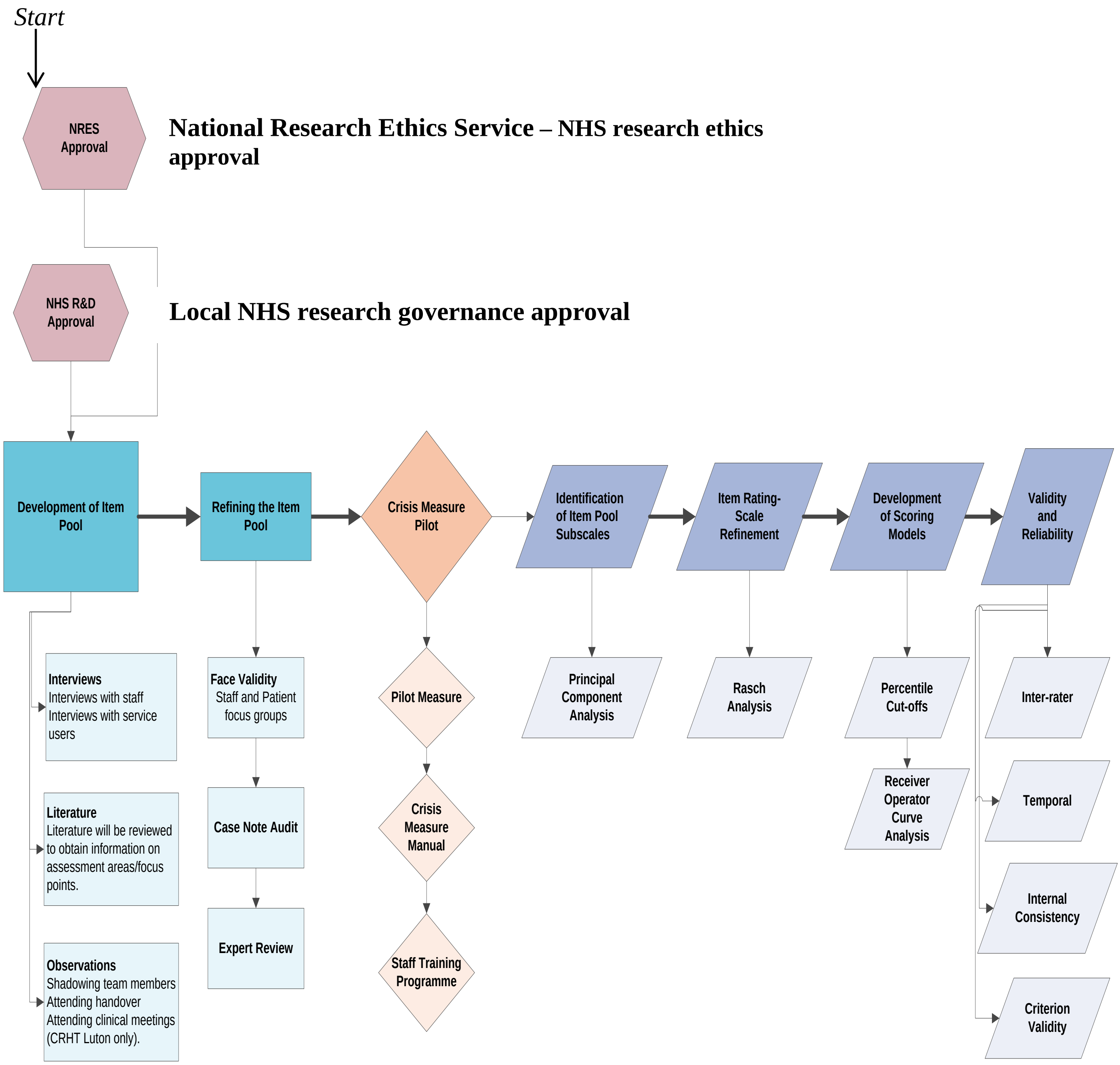
The first phase of the research will concentrate on unravelling the concept of crisis using a qualitative research approach based on obtaining information about the construct from a number of sources including the literature, experts in the field (staff and patients of the CRHTs) and from observation. The information obtained in the first research phase aims to develop an item pool which comprehensively describes the construct of crisis and the item rating-scale to measure it.

The second phase of the research focuses on the development and refinement of a crisis measure using quantitative techniques embedded in both CTT and Rasch analysis. The final stage (phase 3) will provide evidence for the quality of the resulting measure through assessment of the validity and reliability. Figure 1.8 below provides an overview map of the research plan.

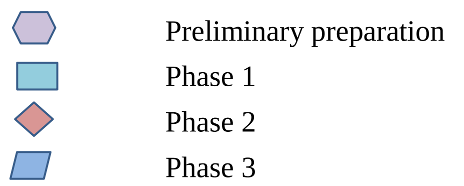
1.4.1 Defining the Measurement Focus

The tool developed through this research aims to support the assessment and treatment decisions of the services offered by CRHT teams to individuals experiencing acute mental health crisis. As defined by the Department of Health (2001) the CRHT teams aim to support individuals ''with an acute psychiatric crisis of such severity that, without the involvement of [CR/HT], hospitalisation would be necessary'' (p. 11). Therefore the tool developed through this research aims to measure the level of treatment required by a patient from the acute services of the CRHT and inpatient care. This tool will primarily aim to have the ability to reliably differentiate between those requiring and not requiring treatment from acute services and secondly to differentiate between the levels of CRHT treatment required. It is expected, based on the previous research of Tobitt & Kamboj (2011, Chapter 1, section 1.0.2.4) that the crisis measurement tool developed will take into consideration functional disruption (including ability to cope, to care for self and sense of mental control), risk of harm (including to self and to others) and additional support required (including current resources and support networks).

Figure 1.8 Overview of Research Plan



Key:



Chapter 2

Discovering the Crisis Construct

Developing the Item Pool and Rating Scale

2.1 Background

It is clear from the literature review that there are no valid and reliable psychometric tools currently available to adequately appraise the needs of people referred to CRHT (Chapter 1, sec 1.0.9). It appears that the main obstacle is the lack of investigation into the concept of crisis itself. In the Bonynge & Thurber study that looked to develop a crisis measure, five measurement items were identified (Chapter 1, section 1.0.9). In their research, there was a distinct lack of evidence to defend the five item's ability to comprehensively represent the concept of crisis. Content saturation or exhaustion is a crucial step in measurement development. This identifies and understands the elements of the concept/construct to be measured (Messick, 1989) and it is clear that the measure developed by Bonynge & Thurber did not achieve this.

There is a wealth of literature describing crisis from observations of professionals and academics in the field (Aguilera, 1998; Hoff, 2001; Johnson, Needle, Bindman & Thornicroft 2008; Roberts, 2002) but to date there is only one structured study that attempts to identify the substance of crisis in a real world clinical setting (Tobitt & Kamboj, 2011) and this only assessed the perspective of the mental health professional

In recognition of this, the first step in this research was to investigate the substance of crisis assessment, to understand fully the concept of crisis from the perspective of those for whom this measure was being developed i.e. mental health professionals and individuals who have experienced crisis. This provided a selective pool of items to represent the possible infinite population of items that exist. The hope is that the items identified would adequately describe the substance of crisis for the development of an item pool from which the final measure would emerge.

The process for identifying the item pool for this measure required an approach consistent with ensuring sound construct validity. A construct validation process aims to identify the construct under investigation (Cronbach & Meehl, 1955). Consideration is given to whether a single construct of crisis is being assessed or a number of independent constructs that represent an underlying key theme. This process identifies items that are representative of the elements of the construct, in this case acute mental health crisis. To appreciate the make-up of crisis assessment requires content coverage or authenticity to be achieved. Messick (1995) and Clark & Watson (1995) both described the significance of the item pool development phase and the potential impact this could have on the quality of the measure when significant omissions in the item pool are made. Therefore, particular attention was paid to developing the item pool and rating scale for this measure, the evidence for which will be outlined in this chapter.

It is important to identify items that provide a comprehensive representation of the construct including the substance and *levels* of the construct. The level of the construct is the amount of the construct present, which is often demonstrated through the format in which the item response is presented. This could be a dichotomous response set (that is a choice of two responses for example yes/no, right/wrong) or a polytomous rating scale (3 or more categories on a scale). The rating scale is there to provide information on the item of focus. Where poor rating scales are used to indicate the level on an item, the information provided is essentially inaccurate, making the item redundant in terms of the usefulness of the information it provides. The implication is that the method for developing the individual item rating scales is equally as important as the method for developing the items themselves. Traditionally the rating scale has not been given as much attention as the development of the item pool. For example, Goldberg and Williams (1991) described in detail the item selection criteria for the General Health Questionnaire but did not describe how the item rating scales were developed. Similarly, the Bonynge & Thurber (2008) study to develop a crisis measure outlined a 5 point likert-style scale ranging from 0 (not present) to 4 (extreme) with no explanation as to *how* or *why* this scale in particular was developed to measure the items.

This chapter focuses on the development of the item pool and the item rating-scale for this measure. This includes an exploration of the scale perspective in terms of a subjective or objective perspective, the method and process for identifying the item pool using CRHT staff and previous patient interviews and focus groups, development of the item rating-scale and the scoring approach used. Finally, the pilot for the first version of the crisis measure developed is described.

2.2 Scale Perspective

It is helpful at the early stages of scale development to decide on the measure's perspective i.e. is the measure going to be an objective or subjective measure. The term 'objective measure' describes a measure that records the observations of behaviour, mood and mental functions of others compared to a 'subjective' measure that records observations of the self. This decision guides the development of items and in particular the development of the item rating scale.

To make a decision regarding the perspective of measurement, it is useful to appraise both the subjective and objective approaches to compare their strengths and weaknesses in the context it will be used. To place this decision in context, based on the literature review completed (Chapter 1) and current clinical practice in the Bedford and Luton CRHTs, it is clear that the assessment information which guides treatment decisions is particularly complex. Information is obtained relating to a large symptom spectrum and associated areas of interest including social issues, protective factors, engagement factors and health difficulties. Information is obtained from a wide range of sources including the G.P., significant others, the patient themselves, other mental health services and the police to name a few.

Crisis is known to impact on a number of key psychological areas including the influence it has on problem solving ability, ability to manage challenges to life goals (often observed as poor concentration) and difficulties with cognitive processing. Patients experiencing crisis often require support to attend to the most basic of needs such as those identified on the first level of Maslow's hierarchy (1943) but also with more complex aspects which require helpful and adaptive coping skills which have been compromised during the crisis experience (Roberts, 2000). Subjective self-report measures can be really useful tools as they essentially cut out the 'middle man'. The experience of the individual can be accessed directly in a more personal way which is both an advantage and a criticism of this approach. However, considering the complexity of the crisis presentation, the significant challenge posed to patients' coping ability, concentration, cognitive processing and the acute nature of distress experienced by patients, it would be unrealistic and unfair, if not unethical to expect a person experiencing crisis to complete a complex measure to accurately reflect their difficulties during this time. In circumstances where the individual was able to complete the measure, the reliability of the outcomes would be questionable.

The crisis measure developed through this research will accompany the patient on their journey through crisis. It will aim to give an indication of crisis level at the beginning of the journey, highlighting areas of difficulty for treatment focus, as well as tracking their progress along the path to recovery. It is important to maintain a consistent approach to measuring crisis and therefore if measurement is to be subjective, e.g. completed by the patient, it will have to remain so throughout their journey. With the distinct possibility that patients will either be unable or not wish to complete such a complex measure during this particularly challenging period in their lives it is sensible to consider more objective approaches to measurement.

Objective measures have the advantage of being completed by someone other than the patient themselves. This may help to provide outcomes that are less emotively and experientially fuelled. However, it may be argued that no one is in a better position to describe their own experience than the person/patient themselves. In the context of CRHT work it is important to recognise that an ‘objective’ measure would be completed by an experienced mental health professional who has assessed a number of individuals experiencing crisis before. They will therefore have this basis of experience and expertise to draw on.

There are advantages and disadvantages to both approaches. The deciding factor came down to the practicalities of ensuring that all patients had completed measures at a number of points along their crisis journey for the purposes of assessment, informing treatment decision making and monitoring. Based on the understanding that it is less likely for the individual in crisis themselves to complete the measure, and that measure completion may potentially increase distress especially in the acute phase of crisis, the decision was taken to develop an objective crisis measure to be completed by the mental health worker. This view was corroborated by patients, mental health professionals and the crisis literature.

2.3 Methods for Item Identification

As discussed in section 2.1, the main concern here is to ensure content coverage by exhausting the construct of crisis (Messick, 1989). Two main approaches for identifying information for the item pool were utilised in this research:

1. Interviews with staff and previous patients from both the Bedford and Luton CRHT teams.
2. A comprehensive review of the literature (as outlined in Chapter 1) to identify the current understanding and theory relating to the concept of crisis. Systematic and comprehensive search of the literature was completed using search engines including PsychInfo, Science Direct and Medline.

Both of these approaches were utilised for the purposes of identifying the item pool in this research. The literature review is outlined in Chapter 1 and the staff and patient interviews are outlined below in section 2.4.

2.4 Staff and Patient Interviews

The CRHT teams in Bedford and Luton provide a community home treatment service that were set up to act as a gatekeeper to inpatient admission. This was to support patients experiencing crisis to be treated in their home environment where possible. The CRHT teams are multidisciplinary, made up of support workers, nurses, psychiatrists, psychologists, social workers, and occupational therapists.

As outlined in the operational policy for Bedford acute and crisis services (2009) the team was set up to provide acute and emergency services in the community. Their objectives were to monitor, provide home treatment and develop care plans for patients experiencing crisis. The team receives referrals from health professionals or self-referral through Accident and Emergency departments. The CRHTs are an adult mental health service and therefore they only treat individuals who are aged 17+ and reside within their catchment area i.e. Bedfordshire.

Interviews were completed with staff and previous patients of the Bedford and Luton CRHTs in recognition of their expertise by experience, placing them in the ideal position to comprehensively describe the construct of crisis. Ethical approval was obtained from Hertfordshire Local Research Ethics Committee; reference number 08/H0309/5 (Appendix 1) and research governance approval was obtained from the Bedfordshire and Luton Mental Health and Social Care Partnership Trust (BLPT) (reference number RGAG-2007-02/12) before the recruitment of participants commenced.

2.4.1 Sample

Number of participants

The approach for identifying the item pool was based on grounded theory (Charmaz, 2003) by adopting the concept that information develops from data rather than being imposed on to it. This approach guided the interview phase of this research. It is underpinned by the understanding that themes and theory are developed from data and not imposed on the data by the researcher (Stern, 1985). The aim in identifying the item pool was to ensure that the available information has been exhausted and the construct itself well represented. Data collection should therefore continue until the evolving themes have been exhausted. Based on this understanding, the analysis of the data itself dictated the number of participants recruited to this research based on the aim of exhausting the themes and as a consequence the crisis construct (for example, Chiovitti & Piran, 2003; Messick, 1989). Strauss and Corbin (1990) refer to theoretical saturation as the point at which no new themes are being identified. Utilising the approach of Strauss and Corbin, the recruitment of participants was ceased once it was agreed by the research team that there was evidence of repetition of previously obtained data with little new data being elicited. For the purposes of this research, the aim was to identify items for the measure's item pool rather than themes but the same logic and approach was applied and participant recruitment ceased once repetition of items was shown and no new data was being obtained.

Staff Recruitment

The staff population was accessed through the Bedford and Luton CRHT teams. All CRHT team members who met the inclusion and exclusion criteria were invited to participate in the research interviews.

- Inclusion Criteria
    - Employed member of staff of the CRHT
    - At least 6 months experience working with the CRHT
- Exclusion Criteria
    - Not employed by the Bedfordshire and Luton Partnership NHS Trust for the CRHT
    - Less than 6 months experience working in the CRHT

Participant Information Sheets were sent by email as part of an invitation to participate in a research interview (Appendix 2). The Participant Information Sheets outlined the overall aims of the research and what participation in the research interviews would involve. Participation was voluntary and interest to participate in the interviews was shown by responding to the email.

Team members who agreed to participate attended interviews that lasted between 1 to 2 hours. Informed consent and basic demographic information were collected before the start of the voice recorded interview. The response rate was reasonable with 12 participants responding out of a possible 44 (27.3%). The sample that responded to the invitation to participate were representative of all the professional areas of the multidisciplinary CRHT teams including nursing, psychiatry, psychology, social workers and support workers. There was also good representation of length of employment with the CRHT (years of employment ranging from 0.5-4 years) in the context that the CRHTs had only been running for 3-4 years at the time of recruitment, gender (7 males and 5 females) and age (age range from 27 to 60 years of age). Participation was voluntary and restricted to the two CRHT teams who agreed take part in the research. Therefore, it is important to consider the representativeness of the outcomes in light of the 335 crisis teams operating in the NHS in England along with the varied demographics of those working for them. However, this is a limitation of all research that uses a sample to represent a population, therefore the key was to take this into consideration throughout the process of the research and in the interpretation of the findings.

Patient Recruitment

All patients who met the inclusion and exclusion criteria who had been under the service of the CRHT between February 2007 and January 2008 were invited by letter to participate in a research interview.

- Inclusion Criteria
  - Previously been a patient of the CRHT
  - Discharged from the CRHT for at least 2 months
  - CRHT team agreed the patient was appropriate for participation
- Exclusion Criteria
  - Current acute mental health crisis
  - Current support from the CRHT
  - CRHT expressed significant concerns regarding the patient's participation
  - Discharged less than 2 months prior to research participation
  - Discharged more than 12 months prior to research participation.

A Participant Information Sheet was attached to the invitation letter (Appendix 3). Previous patients of the CRHT who agreed to participate attended a 1-2 hour interview. Informed consent and demographic information was collected before the start of the voice recorded interview. The response rate was approximately 4.5% of those invited to participate. Four males and 11 females agreed to participate with an age range from 31-60 years of age. It is important to acknowledge the limitations of the representativeness of this sample. The response rate of 4.5% was taken from a sample of the possible population of individuals who have experienced a mental health crisis. It would be interesting to consider the reasons why the response rate was not higher and why these particular individuals chose to respond and participate. Although this was not within the scope of this research, the research was carried out mindful of this limitation and the possible consequences e.g. the potential for missing important information and therefore not achieving content coverage. However, this is a limitation for all research that uses samples of a population.

2.4.2 Overview of the Staff and Patient Interviews

Grounded Theory provides a clear outline for structuring semi-structured interviews (Smith, 2003). This approach guided the development of the interview outline which was carried out with participants. The grounded theory approach was not adopted in a formal manner but simply used to guide the development of the semi-structured interview. Staff participants were asked to talk about crisis assessment and patient participants were asked to talk about their experience of crisis in terms of three phases which reflect the treatment phases used in practice by the CRHTs. The CRHT use a traffic light system of Red, Amber and Green treatment status to indicate acute, moderate and low crisis states:

1) Initial acute crisis phase – this is the acute phase of crisis where an individual is demonstrating a significant breakdown in their coping ability. This generally requires regular support from the CRHT normally once a day but can be up to three times a day. The acute phase of crisis is categorised as Red on the traffic light system (Red/High treatment status).

2) Stabilisation phase – this is the phase of treatment where the patient starts to show signs of recovery in their coping ability, moving away from the acute crisis phase and showing signs of stabilisation. This phase of crisis is categorised as Amber on the traffic light system (Amber/Moderate treatment status).

3) Recovery phase – this phase is where the patient is demonstrating good recovery and utilisation of helpful coping strategies, showing movement towards independence and away from dependence on the support of the CRHT. This phase of crisis is categorised as Green on the traffic light system (Green/Low treatment status).

Based on the three phases of crisis outlined above, two types of questioning were used to elicit information about the assessment and experience of crisis through the staff and patient interviews:

1) Basic open ended questions for example: ‘Tell me about the first phase of crisis’

2) further information was elicited through a technique similar to Kelly’s (1955) method of triadic questioning, for example: ‘Can you think of ways where these phases that you described are alike, yet different from this third phase?”. This approach was chosen to elicit information regarding the

underlying crisis construct but in addition to try and identify an appropriate rating scale on which to measure the construct i.e. to obtain information on how severe or significant the crisis state is as indicated by the CRHT treatment status of Red, Amber or Green.

Each interview generated an audio recording of the interview and a flip chart map of the main themes discussed (example of interview flipchart record, Appendix 4).

2.4.3 Interview Outline

The aim of the interviews was to draw out as much information as possible from participants regarding their experiences of crisis (patient interview focus) and crisis assessment (staff interview focus).

At the beginning of each interview session the main points of the Participant Information Sheet were summarised and the opportunity for questions given. This summary detailed the reasons for completing the research, the voluntary nature of their participation, confidentiality and data use and storage. The interview was semi-structured and audio recordings were made of all interviews as consented to by all of the participants.

All staff participants were asked to bring an example patient case for each of the crisis treatment phases. The purpose of this was to support the staff participant to think about the phases of treatment with real life examples. For some of the interviewees, reference to these examples was not necessary whilst for others this acted as a useful prompting tool.

The first set of questions looked directly at each crisis phase, for example '*I would like you to think about the first phase of crisis for a moment. Can you tell me about the factors you identify as most relevant to this phase?*' This was used to ask about each phase of crisis. This was a semi-structured interview format and therefore, although each participant had the same initial question asked for each of the three phases of crisis, the follow-up/prompt questions varied depending on the initial response given. For the majority of participants this simply required prompts to provide further information for example '*can you tell me more about that.*' As the participant talked about the factors they would take into consideration or had experienced, the main ideas or themes of their responses were noted down on flip chart paper. Once they had exhausted their ideas, the main factors were reviewed with the aim of confirming the information already obtained and to prompt further discussion. For the majority of participants, this approach to questioning elicited a wealth of information and exhausted their approach or experience of crisis assessment or crisis.

A second set of questions were designed to obtain further information that provide guidance toward an appropriate rating scale for gathering information on the items. The Triadic Comparison Method (Kelly, 1955) for questioning was used for this purpose, for example, '*How does the first phase of crisis differ from the last two?'* The comparison of one crisis phase against the other two crisis phases was useful for participants who found talking about the phases with open questions more difficult.

The final part of the interview asked the participant to underline the factors they felt were most important to each of the phases. This could be one, several or none of the factors.

On average the interviews took an hour and a half to two hours. For one of the more experienced nurses a second interview was set up to discuss in more detail some of the factors the nurse had underlined as being the most important. The first interview took 2 hours and therefore a second interview was arranged in recognition of the possible impact of interview fatigue on the participant and the consideration of information quality should the interview have been continued.

2.5 Utilising the Interview Data - Preparing the Item Pool

Throughout the interview the key items were noted down on flipchart paper and reviewed at several points during the process (Appendix 4). At the end of the interview the items were confirmed with the participant and all items were initially entered into an excel spread-sheet. A total of 645 items were noted down from the interviews, the majority of these mirroring what had been noted from other participant interviews. Items were sorted alphabetically in excel and items that were the same were merged together.

2.5.1 Number of Items Retained in the Item Pool

As indicated in the literature review (Chapter 1, sec 1.1.6.3), no amount of statistical analyses can make up for omissions made in the item pool development phase (Clark & Watson, 1995). Decisions regarding item retention are delicate, a fine balance between providing a comprehensive representation of the construct and developing a measure that is practical. The items to retain for the item pool were considered in light of practical, statistical and theoretical perspectives. There are a number of practical considerations such as length of time to complete the scale and rater bias resulting from when there are too many or too few items. Too many items and the rater may experience fatigue and response pattern bias (Anastasi, 1976), too few items and the content and construct reliability and validity may be compromised (Kenny, 1979). There are also a number of statistical considerations, for example, Cook, Hepworth, Wall, & Warr, (1981) showed that adequate internal consistency could be obtained with as little as 3 items but for a more complex construct more items/variables would be necessary to represent the construct adequately. As more items are added the impact of individual items on the reliability of the tool decreases (Carmines & Zeller, 1979). A theoretical consideration is that items included in the scale must provide content coverage to ensure that the construct has been adequately represented. These practical, statistical and theoretical considerations were used for decision making regarding item retention for this crisis measure.

Items were removed where similarities, clinical overlap or repetitions were observed, as agreed by the research team. A list of 133 items was obtained and this was considered acceptable in terms of meeting the practical and theoretical considerations outlined above.

To confirm the item pool obtained from the interviews and literature review, the item list was presented to two focus groups– one consisting of CRHT staff and the other consisting of previous CRHT patients. All staff and patient participants who responded to the original participation request, regardless of whether or not they took part in the interview stage, were invited to attend the focus groups. Consent was obtained from patients for confidentiality. A total of 14 participants took part in the focus groups. Five members of the Bedford CRHT team and 3 members of the Luton CRHT team participated in the staff focus group. Six patients participated in the patient participant focus groups.

Participants were asked to consider the list of items and discuss:

1. How clinically useful the items were on a scale of usefulness to uselessness

2. Frequency of use (staff only)

3. Further items to be added to the item list.

Items that were deemed useful (i.e. rating on the ‘useful’ side of the likert-type rating scale) by >80% of participants were kept in the final item list in line with the thresholds set out by Trochim (1989a) for concept mapping. It was found that all of the items were considered to be useful for CRHT assessment. Three of the items (libido, feelings of uselessness and childhood development) were marked by more than 2 participants as being less useful and were therefore considered for removal from the item pool. However, it was found that ‘libido’ and ‘childhood development’ were regularly rated as part of on-going assessment as indicated by the staff participants and therefore these were retained in the item pool. ‘Uselessness’ was not rated regularly and was also found to be less useful so it was removed from the item pool. Two further items were added to the list – ‘weight loss’ and ‘physical health problems’ as identified by the focus group members. This resulted in a total item list of 134 items.

For the purposes of piloting the items in the CRHT teams it was agreed in the focus groups that grouping items together by common theme would make it easier for CRHT team staff to conceptualise the measure and therefore to complete it. Items were grouped together using a basic approach of concept mapping (Trochim, 1989a) with input from three Psychologists. 13 groups of items were identified as outlined in Table 2.0 by grouping items together that appeared to relate to an underlying common clinical area. With the addition of item group titles, the item list came to 147 items.

Table 2.0 - The 13 Item Groups of the Pilot Crisis Measure

Number	Item Group Title
1	Overall Behaviour
2	Overall Thought Content and Clarity
3	Overall Feelings/Affect
4	Overall Risk
5	General Support, Buffers and Protective Factors
6	Overall Awareness & Psychological Mindedness/Crisis Understanding
7	Overall Crisis Presentation
8	Overall Vulnerability
9	Overall Coping
10	Overall General Wellbeing
11	Overall Historical Indicators
12	Overall Psychotic Symptoms
13	Overall Acceptance of Support

Table 2.0: Outlining the item groupings in which items were clustered together based on clinical similarity for the purposes of piloting the crisis measure.

2.5.2 Small Scale Case Note Audit

The 147 items identified for the crisis measure were developed and confirmed from information provided by the CRHT mental health professionals and individuals who had experienced crisis and had previously received CRHT support. Therefore, it would be expected that the majority of these items, deemed useful by over >80% of both staff and patient focus groups, would be present in the clinical notes. To validate the item pool for content coverage a sample of CRHT assessments from both the Bedford and Luton CRHT teams were compared to the item pool to see:

1. Whether the item list contained all of the items assessed by the CRHT according to the assessment notes.
2. The percentage of items identified through this research that had been documented in the clinical notes of the CRHT teams.

Half of the caseload notes were taken from each CRHT team at random. Fourteen files from the Bedford CRHT and ten files from the Luton team were reviewed and compared to the item list from this research. An additional item was identified from the initial review of the assessment notes and added to the item pool bringing the total to 148 items.

On average significantly fewer items were being documented in the clinical notes compared to those outlined in the item list. Bearing in mind that these items had been developed through interviews with CRHT staff involvement and the final item pool agreed with focus groups, it is surprising that assessment notes had so few of the identified items documented.

Twenty seven (18.2%) of the items were not included in any of the assessment notes reviewed. This included items such as fluctuations in behaviour, historical coping strategies, goals, ability to control/manage thoughts, complexity of presentation, ability to manage symptoms and crisis state. The 2 items that were evident in all notes were grouping items – Overall Psychotic Symptoms and Feelings/Affect. It was evident from the file review that content coverage of acute mental health crisis was not being achieved through the traditional note taking approach that was in use by both teams. This provides further support for the development of this crisis measure and is discussed further in the Summary section of this chapter (section 2.13).

As a result of this process and a final focus group review of the item pool, 5 further items were extracted. A final item pool of 143 items was agreed for the first scale pilot.

2.6 Development of the Rating Scale

This section focuses on the development of the rating scale used to represent the item level. There are two main methods for rating items, either with a dichotomous response set or polytomous scale (Chapter 1, section 1.1.6.4). A Likert-style (1932) scale is a scale capable of differentiating between different types of response and to which values can be assigned for analysis. The statistical reliability of Likert-type scales has been shown to increase with an increase in the number of categories up to 5 (Lissitz & Green,1975). However, in addition to statistical reliability, it is important to develop a scale that has clinical credibility, makes sense to the assessor and can be easily communicated to the patient. Therefore, the validity of the rating scale is based on statistical and theoretical as well as clinical and logical reasoning.

2.6.1 The Rating Scale

The rating scale developed through this research was initially identified through the staff and patient interviews. There were a number of common themes identified from the interviews around the impact of crisis on the individual's ability to cope, ability to make healthy decisions, predictability and ability to stay safe. The main themes are described here with excerpts from the research interviews.

Ability to cope and manage

"Somebody may still have fleeting thoughts [of suicide] and may have always had fleeting thoughts but I would be looking at if it has reduced and have they managed those thoughts." (Participant 1, p.10, sec.123 pp.323-325)

One of the repeating themes across both the staff and patient interviews was this idea around abililty to manage and cope. In the above extract, the participant expresses that it is not only the intensity of symptoms but the ability of the individual to manage and cope with those symptoms.

Impact of crisis on decision making ability

"…who has relapsed to a point where her alterations in perceptions, thoughts are so disturbed that she is not able to make that judgement for herself" (Participant 2, p.2, sec.2. pp58-59).

Another common theme across the interviews was around the impact of crisis on a person's cognitive ability and therefore their ability to make healthy decisions. In this example the theme centres on the impact of crisis on the individual's ability to make helpful and healthy decisions.

Impact of crisis on predictability and ability to stay safe.

"I think for me what would make me make an acute admission might be someone who is acutely psychotic, their level of dangerousness, behaviour towards themselves or other people would make me go for an acute admission rather than crisis or home treatment unless it is somebody you have worked with before." (Participant 2, p.3. Sec. 4. pp72-78).

In this example, the decision to make an admission is not purely based on the immediate presentation but on whether or not the patient is someone known to the service previously and therefore is more predictable as a person. There is a strong emphasis on the patient's ability to stay safe in terms of their likeliness to harm themselves or to harm others.

The over-arching theme emerging from the interviews is that the presence of symptoms or crisis state itself contributed only in part to the treatment decision. The main considerations are focused on the ability of the individuals to manage and cope with their difficulties in a helpful way, to make adaptive decisions and to safely manage risks in the community. The more predictable a person is, the easier it is for the team to assess the individual's ability to meet these criteria, the less cause for concern the individual is.

Therefore two patients could score similarly for the *presence* of a symptom or item but one may be considered appropriate for admission and the other considered for home treatment. For example there may be two patients both experiencing the symptom of '*thoughts of suicide'* at similar levels or frequency but the assessment outcome may differ depending on a number of factors, for example protective factors, previous history and ability to cope based on previous contact with services. The symptom itself is only part of the treatment decision considerations. It is a combination of different presenting factors that ultimately determines how much of a cause for concern the symptom is.

It is not simply the *presence* of the symptom that is important but how able that person is to cope and manage, to make helpful decisions and to manage their risks in the community. The focus here, the overarching theme, is on *how concerning the symptom* is for that individual in the context of their presentation.

Crisis assessment takes into consideration how safe it is to treat the person at home and as outlined above this focuses on the risk factors in terms of how well the individual is able to manage and cope with the symptom/item. Balancing factors to the potential risk factors are the protective factors. For example, protective factors may include an individual's support systems, family and networks. A number of the staff participant interviews indicated that the protective factors and support systems available to the individual patient were important aspects for treatment decision making. Two examples are given below.

Example 1 – Protective factors and support systems

"It depends, again on whether the Service User is living alone, whether he has support, whether the person who is supportive at home is at the end of his tether, whether the Service User is compliant with medication, not compliant with medication would mean we would have to supervise and give the medication." Participant 3. p.3, sec.22 pp. 86-89.

Example 2 – Protective factors and support systems

"Or if they are not able to, somebody in the family is able to, without too much problem. Somebody can be a bit depressed but can be cajoled or encouraged to eat and then with our support with medication…." Participant 2, p4. sec.12. pp137-140

CRHT staff commonly assess risk and protective factors to support their decision making. Weighing up these factors, the clinician is able to understand how much of a cause for concern the patient is in relation to home treatment. The assessment must encapsulate the patient's own internal resources and resilience as well as the support structures around them. Therefore the scale developed here would be most useful and informative when it could take into account both risk and protective factors to understand the level of concern for the item being assessed.

2.6.2 Scale Wording

The item rating scale wording is crucial as it summarises what the item scale is attempting to represent. The rating scale aims to capture the change in a person's presentation that signifies the break down in coping associated with crisis. Therefore, the focus is on change with the individual moving away from their natural state of balance and equilibrium as suggested by the associated theories of homeostasis and equilibrium (Chapter 1, section 1.0.4.1). Four different wording options for the scales were developed based on the information obtained from the interviews. The four rating-scale wording options were shared with the focus groups for consideration:

a) Risk – the rating scale would represent risk running from *no risk* to *risk*.

b) Cause for concern – this rating scale would run from *not cause for concern* to *cause for concern*.

c) Barrier to treating at home – this rating scale would run from *barrier to home treatment* to *not barrier for home treatment.*

d) Barrier to treatment – similar to the above wording, this rating scale would specifically relate to CRHT treatment running from *barrier to treatment* to *not a barrier to treatment.*

A unanimous decision was made to use the wording '*cause for concern'* for the scale as it was felt this represented the construct being measured and the language was felt to be less intimidating or threatening for patients as well as easier to communicate.

2.6.3 Rating Scale Format

The rating scale developed was deemed by both focus groups as clinically suitable for rating items in relation to crisis assessment. The rating scale chosen is an 11 point rating scale that takes account of both risk and protective factors with the overarching scale title of '*cause for concern'*. The rating scale style agreed is based on FIT profiling developed by Fletcher (2003) which supports flexible rating of items using either a single score or a range of scores on a polytomous scale. Profiling rating scales that support flexible rating of items using either a single score or a range of scores on a polytomous scale allows far greater flexibility in rating the scale. Moreover it is unnecessary to reverse score on any item as the orientation of the item (negative or positive) becomes neutral with the scale rating itself providing the item orientation. This scale format was chosen due to the clinical strengths it offers and because previous uses of this format have produced impressive reliability outcomes (e.g. Sharma, 2011).

The main consideration for the scale format was to ensure that it assisted the mental health professional to capture the information imperative to understanding the level of concern. The scale of *cause for concern* identified and agreed through the focus groups was chosen because of its ability to capture both the individual's risk and protective factors. The rating scale balances both risk and protective aspects of an item within the same scale reflecting the clinical reality of assessment.

To represent this concept an 11 point scale has been developed to outline this idea of a balance (Figures 2.0 and 2.1). The central point is the 0 or neutral point which represents where there is a balance between the risk and protective factors e.g. both risk and protective factors are present in a 50:50 ratio. One step up the scale (to the left of 0) represents a slightly higher risk to protective factor ratio where as one step down the scale (to the right of 0) represents a slightly higher protective factor to risk ratio.

The left extreme of the scale represents the greatest level of concern or 100% risk. The right extreme of the scale represents the lowest level of concern or 100% protective factors.

Figure 2.0 - The Flexible 'Cause for Concern' Item Rating Scale

5	**4**	**3**	**2**	**1**	**0**	**1**	**2**	**3**	**4**	**5**

Cause for concern **Neutral** **Not cause for concern**

Figure 2.0 outlining the 11 point *Cause for Concern* item rating scale running from 'Not cause for concern', through a balanced 'Neutral' point and to the 'Cause for Concern' end of the rating scale.

To conceptualise the 11 point *cause for concern* rating scale as a balance between risk and protective factors, Figure 2.1 below maps this out in terms of risk and protective factor percentages for each point along the item rating scale. Both the risk and the protective factors have scales running from 0-100%. However, these scales run in opposite directions. The central point of the scale is a balancing point where there is a 50:50 balance between the risk and protective factors. As the scale moves towards the *cause for concern* end of the scale the risk factors start to account for a larger proportion of the ratio compared to the protective factors e.g. the third point on the left hand side of the scale indicates that risk is 80% and protective factors account for 20%. At the extreme end of *cause for concern* the risk factors account for 100% and the protective factors account for 0%:

Figure 2.1 – The Flexible Item Rating Scale with Risk and Protective Scales.

	5	**4**	**3**	**2**	**1**	**0**	**1**	**2**	**3**	**4**	**5**	
Cause for Concern						**Neutral**						**Not cause for concern**
% Risk	100	90	80	70	60	50 %	40	30	20	10	0	**% Risk**
% Protective Factors	0	10	20	30	40	50 %	60	70	80	90	100	**% Protective Factors**

Figure 2.1 outlines how the *Cause for Concern* item rating scale combines the 'Risk' and 'Protective Factors' elements of assessment within the same rating scale using the effect of measurement scales where the factors act to balance each other. This has been depicted here using percentages.

Moving toward the '*Not cause for concern'* end of the scale from the 0/50:50 point indicates that the protective factors start to account for the larger proportion of the ratio, e.g. point 3 on the right hand side of the scale indicates 80% protective factors and 20% risk.

Mental health professionals will be most concerned i.e. a score of 5 for cause for concern, when there is 100% risk with no protective factors. Using the metaphor of a pair of scales, the scales will only have weight on the risk side, therefore weighing down on the left hand side of the scales. In contrast a mental health professional will be least concerned (score of 5 for no cause for concern) when there are no risk factors and 100% protective factors. In this scenario the right hand side of the scale will be heaviest. An example of how this scale would be used is outlined in Figure 2.2 below.

A protective factor could also indicate that there is *not cause for concern* when the item is shown to have not occurred previously e.g. if a patient has not attempted suicide previously, the item 'previous attempt of suicide' would be scored as a 5 on the *not cause for concern* end of the scale thereby acting as a protective factor. This was agreed as a suitable scoring method based on the evidence that previous behaviour is the best predictor of future behaviour (Ajzen & Fishbein, 1980).

Figure 2.2 - The Flexible Item Rating Scale with Examples of High, Medium and Low Risk Scores

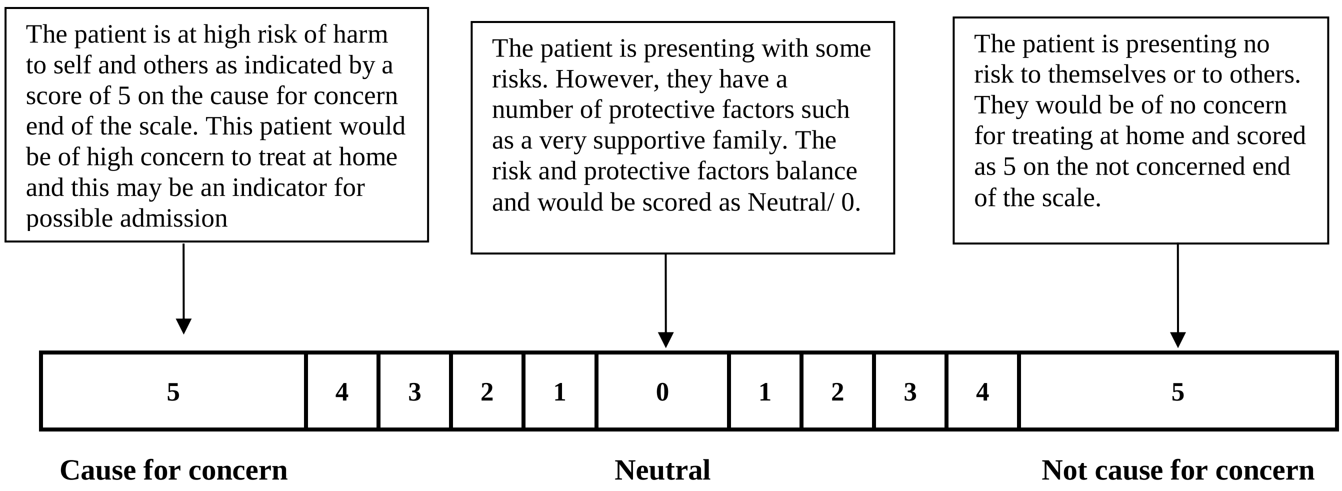


Figure 2.2 outlining three examples of how patients appropriate for different points along the rating scale would be scored based on the *Cause for Concern* flexible item rating scale.

2.7 Rating System

The flexible scoring system adopted for this rating scale recognises that many mental health difficulties are changeable either over time or between situations. The difficulty with a single score is that it is not always possible to decide on one fixed score. It is not unusual for assessors to score either across multiple scores or between two categories on a scale even when this is not given as an option. Use of a flexible rating scale, whereby assessors are given the freedom to use a range, supports the assessor to capture this information. This fits more with the idea of using a continuous scale whereby an assessor can place an individual over a spectrum rather than onto fixed categories. For further explanation the hypothetical example of shyness is used (Figure 2.3 and 2.4). If an assessor is to rate how shy an individual is, it may be shown that the individual is shy in all situations in which case their shyness is stable. In this situation it could be scored using one number, for example (score indicated with an X):

Figure 2.3- Single Point Rating

5	**4**	**X**	**2**	**1**	**0**	**1**	**2**	**3**	**4**	**5**

Cause for concern **Neutral** **Not cause for concern**

Figure 2.3: item rating scale example of a single point rating for the hypothetical item of 'shyness'.

However, a person's presentation may not always be stable and therefore their presentation may differ across situations. Using the example of shyness again, an individual may be confident and talkative when with close family but shy with others who are less familiar. Therefore this person's *cause for concern* would differ across situations (figure 2.4). This would be scored in a manner to represent the range of behaviour from the 'best case scenario' (when shyness is least in the company of close family) to the 'worst case scenario' (when shyness is at its highest e.g. when with strangers):

Figure 2.4 – Range Score Rating

5	**4**		**2**	**1**	**0**	**1**	**2**	**3**		**5**

Cause for concern **Neutral** **Not cause for concern**

Figure 2.4: item rating scale example of a range score rating for the hypothetical item of 'shyness'

2.8 Scoring the Flexible Rating Scale

For the purposes of providing data for analysis and for providing scores for outcome interpretation the range score is summarised by a single score. This is done by using the mid-point of the range (Figure 2.5). Where the mid-point falls between two scores, the higher score (the one closest to the *cause for concern* end of the spectrum) is taken to ensure that the risk factors have been captured.

Figure 2.5- Scoring the Mid-Point.

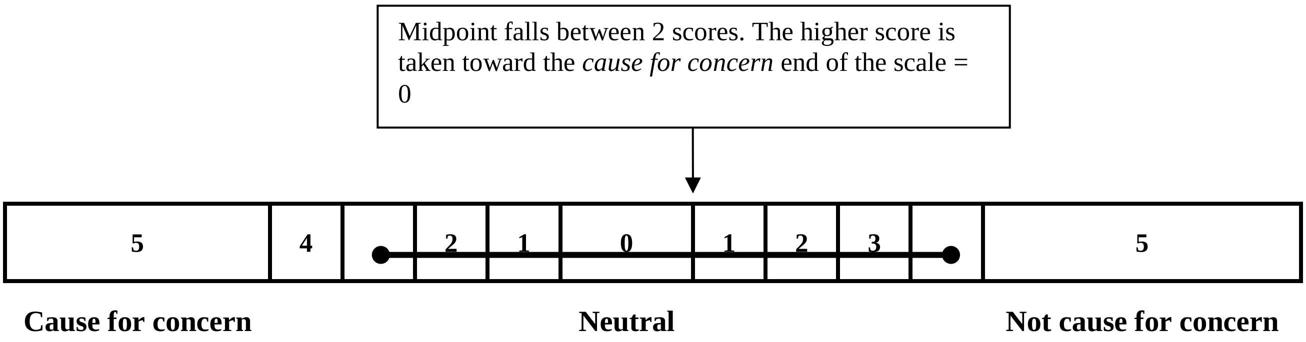


Figure 2.5: item rating scale example of scoring the mid-point of a range rating on the flexible scoring scale.

The scoring system used for the purposes of outcome interpretation and data analysis runs from 0 to 10 from the *not cause for concern* end of the rating-scale (point 5 = score of 0) to the *cause for concern* end of the rating scale (point 5 = score of 10) which makes this an 11 point scale. For example, a person receiving a rating of 0 at the midpoint of the rating scale would receive a score of 5 for the purposes of outcome and data analysis.

2.9 The Use of Not Applicable (N/A)

It was decided that the use of a ‘Not Applicable’ (N/A) category was not helpful for this measure. The scale is trying to capture the elements of crisis and to identify how concerned the professional should be about the patient’s ability to cope safely at home. Where items may be viewed as *not applicable* often indicates a protective factor. For example, it may be thought appropriate to score the item ‘Regret of suicide attempt’ as N/A when that individual has not previously attempted suicide. However, the very fact that the individual has not previously attempted suicide in itself acts as a protective factor based on the evidence that suggests previous history is one of the strongest predictors of future behaviour (Ouellette & Wood, 1998; Trandis, 1977, 1979). Therefore, no previous history of suicide attempt presents as a significant protective factor and would reduce the overall *cause for concern* which would be rated as 5 for the *not cause for concern* end of the scale which indicates a score of 0 for that item. This would be more meaningful than indicating the item as ‘Not

Applicable' which would not capture the true information provided by no previous history of suicide attempts.

2.10 Timescale for Completion

CRHT team members may complete up to 5 patient visits in one shift. As a consequence, it can be difficult to accurately recall information if too much time is left between the visit and completion of the paperwork. It was agreed with the CRHT teams that it would not be appropriate to complete the scale whilst with the patient as this would act as a barrier to the therapeutic relationship and may increase the patient's anxiety levels or demand levels of concentration that a person in crisis is not able to provide. Therefore it was recommended that the scale was completed immediately following the visit for the pilot of the measure. In terms of practical implications and future measurement development, it may be useful to examine the potential for using portable electronic devices on which the measure could be scored whilst the mental health professional is out in the community.

2.11 Crisis Measure Pilot

2.11.1 Sample

Nunnally (1978) recommended that ten datasets should be completed per variable. Kass and Tinsely (1979) recommended between 5 and 10 times the number of completed datasets as variables up to 300 where it is believed that findings become stable despite adding further datasets. This was confirmed more recently by Tabachnick and Fidell (2007). Categorising into poor, good and excellent categories, Comrey and Lee (1992) classed 100 as poor, 300 as good and 1000 as excellent. Therefore, for the purposes of this research a minimum of 300 completed scales was the minimum dataset to obtain for analysis. This was a realistic and achievable goal with access to two CRHT teams made up of an average of 40 staff members within each team and with an estimated average caseload of 30 patients at any one time.

Following the pilot of the crisis measure, 385 measures were completed by the CRHT staff and the data entered into SPSS (PASW statistics 18 software, 2007) and RUMM2030 (2010) software for analysis.

2.11.2 Development of the Training Manual

A training manual was developed to support the staff training and to provide a reference tool for staff following training. The manual developed outlined the basic principles of the crisis measure including a description of the rating approach for the single point and flexible rating, the cause for concern rating-scale and a definition list to describe the items. The definition list was developed to not only provide a definition of each item but to provide prompts for completing the scale. This was developed using definitions from The Oxford Dictionary of Psychology, the Diagnostic and Statistical Manual of Mental Disorders (IV), Psychiatry – An evidence based text (Puri & Treasden, 2010) and clinical experience from the CRHTs. The manual served as an operational tool for the pilot and was not changed throughout the duration. Following completion, the manual will be updated to include information regarding the outcomes of this research, in particular, information concerning the validity and reliability of the measure.

2.11.3 Access to CRHT Teams and Support with the Implementation of the Crisis Measure

In order for the pilot to be integrated into standard routine practice by the teams, it was crucial to obtain top level 'buy in' as this was a major change to staff practice. Liaison with CRHT management and clinical leads helped to obtain ground level support by encouraging and motivating team members to complete the measure as part of routine practice. Supervision structures held in the CRHTs offered forums for staff to seek support with scale completion and for supervisors to emphasise its integration into practice. Due to the nature of implementing change in the NHS it was helpful to have the researcher as a point of contact for the CRHT teams on a regular basis. Therefore it was part of the researcher's role to support patient visits, assessments, interventions and regularly attended handovers.

2.11.4 Crisis Measure Training

The first version of the crisis measure was piloted with the Bedford and Luton CRHT teams, implemented as part of standard practice. All CRHT staff were trained to complete the pilot crisis measure. The training programme approach evolved over the course of the research as a result of feedback received from the CRHT team members and difficulties the team shared regarding their understanding and ease of completion of the measure. There was feedback from the observations of clinical leads through the supervision structure and by the researcher when supporting team assessments. The first phase of training included a large group training session over approximately 2 hours with no further follow up. Due to the complexity of the crisis presentation and the new flexible rating system utilised by the scale it was soon apparent that team members required further training and more practical support in terms of directly applying the scale with support in practice. The training programme was developed further to reduce the group training down to small groups of up to 3 staff members with an initial 2 hour training session followed by a shadowing session. The shadowing session involved the researcher/trainer going out on visits with CRHT mental health staff to support with assessment and subsequent scale completion. Following this, supervision sessions were set up on a needs basis. The main area of difficulty was generally related to understanding the concept of the scale and the flexible scoring system. However, with adequate support and supervision this did not take more than one extra 20 minute session to clarify. It is recognised that inter-rater reliability depends on consistency of item rating between mental health professionals rating the same patient (Field, 2009). Therefore, difficulties relating to how the individual mental health professional understands and uses the rating system on the measure will be reflected in the outcomes of the inter-rater reliability analysis. Due to the difficulties in measure comprehension at the beginning of the training programme, subsequent changes in the training approach, the length of the scale, complexity of crisis presentation and the implementation of the new flexible scoring system it was expected that these inconsistencies would be reflected in the outcomes of the inter-rater reliability analysis (Chapter 8, section 8.3.3).

Both teams work from a multidisciplinary model with support workers, nurses, psychiatrists, psychologists and social workers who recognise the experience of acute mental health crisis as affecting the individual as a whole person, their family and the systems around them. To ensure that the scale data collected for the crisis measure represented a multidisciplinary perspective of crisis, all mental health professionals including non-qualified staff, i.e. support workers, were trained to complete the scale. However, in practice it is unlikely that non-qualified staff would hold the level of responsibility or mental health qualification/expertise to complete the measure independently but for the purposes of obtaining a comprehensive data set to describe the construct of crisis assessment it was important for the research to include the perspective of non-qualified staff who undoubtedly contribute to the CRHT team's perspective and subsequent treatment of patients experiencing crisis.

2.12 Data Collection and Storage

Data was collected anonymously by returning completed scales to a research box kept in the CRHT team offices. All participants were assigned reference numbers to allow comparison for inter-rater reliability and similar analysis later in the research. The individual unique reference numbers for staff participants made it possible to identify 43 individuals contributing completed measures to the data set. It is important to acknowledge that the number of measures completed by each assessor varied with some completing just one and others making more substantial data contributions (Appendix 19). This is a limitation of real world research, whereby control over variables such as standardising the number of measures completed by participants is difficult without self-implementing a serious obstacle to data collection. Statistically, this may bias the findings of the study whereby some participant's 'opinions' will have more weight than others simply as a result of the number of measures completed. It is therefore recommended that larger scale pilots are completed on the measure developed through this research to assess the representativeness of these findings. All patients were assigned a reference number to protect their identity. All data was entered into SPSS using the participant reference numbers only, saved to password protected South Essex Partnership NHS University Trust (SEPT) network. All documentation relating to the research was backed up on a password protected memory stick which was kept by the researcher. All completed scales were stored in a locked filing cabinet in the Department of Psychology on SEPT premises. All data was treated and stored in a manner to meet the criteria of the Data Protection Act (1998).

2.13 Summary and Conclusions

These early stages of measurement development are crucial to validity and subsequently the reliability of the resulting measure. The items from the item pool represent only a sample of all the possible items contained within the item population and therefore it was an aim of this research to obtain a sample that could comprehensively represent the substance and continuum of the construct of crisis to ensure that there were no significant item omissions. At the same time as ensuring a comprehensive description of the crisis assessment construct, it was important to consider bias that could result from unnecessarily long or complex measures where fatigue, boredom and guessing may result in invalid and unreliable outcomes. The focus for the development of the item pool and rating scale for this measure was on ensuring comprehensive items of quality to represent the crisis construct. Subsequent analysis of data obtained for the item pool developed in this chapter may help to identify *where* gaps in the item pool lie but there are no statistical techniques available to identify *what* those gaps are. Therefore it was important at this stage to ensure that the potential infinite item population for the crisis assessment construct was sufficiently represented by the sample of items in the item pool obtained through this research. The steps taken to develop the item pool for this measure clearly demonstrate the comprehensive approach necessary to support the development of a representative item pool for the construct of focus and provide a leap forward in the understanding of the underlying construct of crisis assessment when compared to previous attempts to develop representative measures (Bengelsdorf et al, 1984; Bonynge & Thurber, 2008; Myer & Cote, 2006; Myer et al, 1992).

Although the approach adopted here was comprehensive, it is also important to recognise the limitations of the sample used in this research. Firstly, there is response bias to consider in terms of the sample of the population who responded and volunteered to participate compared to those who declined the opportunity. It was acknowledged in section 2.4.1 that only a small percentage of patient participants responded to the invitation to participate in the research interviews. It was hypothesised that this may have been due in part to individual experiences of crisis and not wishing to revisit that experience for the purposes of research but without further investigation the reasons for this low response rate remain unknown. However, this in itself would warrant investigation in its own right and therefore is simply suggested here as an area of focus for future research. This is a problem for all qualitative and quantitative research and should be held in mind when thinking about the construct of crisis assessment identified here. Reassurance is taken from the fact that items/themes were identified that acted as linking threads through all of the interviews for both staff and patients which provides confidence that content saturation was achieved. In terms of research, lack of research buy in or engagement and drop-out rates are research projects within their own rights and therefore will not be addressed in this research but would be an interesting topic for future research studies.

The case note audit comparing the current approach for documenting crisis assessment and the items identified as important to assessment by both CRHT staff and patients indicated that there is a significant gap between what is being documented in current practice and the ideal outlined in the interviews. In addition, there was inconsistency across different sets of assessment notes and therefore what may have been documented in one case file may not have been documented in another, which demonstrates a lack of consistency and standardisation in documenting assessments. Although the reasons for this gap were not directly investigated, the identification of 143 items for the initial item pool may indicate where some of the difficulties lie in terms of how realistic it is to expect CRHT staff, often visiting up to 5 patients a day out in the community, to write up and comment on 143 item areas of assessment. Development of an assessment measure may support more accurate and objective measurement as well as helping to streamline and standardise the assessment process. Although it would not be reasonable to expect crisis workers to document such comprehensive information considering the time restraints placed on the team, it should also be recognised that without comprehensive documentation risks increase. For example, if a change in the patient's presentation occurs over the course of their treatment with the CRHT, it is unlikely to be identified if the previous presentation has not been documented. Accepting that it is unreasonable to expect assessors to document 143 items at each assessment, it is therefore inevitable that gaps in the documentation will occur and as a consequence risks will increase. The development of a standardised assessment measure is therefore a positive step forward in standardising the assessment approach and reducing risk.

The rating scale developed in this research sets this crisis measure apart from not only previously developed crisis measures but all mental health psychometric measures in use today. The crisis measure's rating scale is set apart for two reasons; 1) the rating scale is not measuring the presence of the item but the level of concern the item presents for the patient and 2) the rating scale encapsulates both risk and protective factors within the same scale based on a conceptualisation of a seesaw/balance effect.

As suggested in Chapter 1 (section 1.0.4.1), the concept of equilibrium (or homeostasis) is an important one for assessing crisis. Every individual has their own state of ‘normal’ within which they function. It is movement away from this balance or equilibrium that indicates change and possible crisis. The item rating scale needs to be able to capture this individualistic approach to assessing crisis, not simply documenting the presence of symptoms/items but assessing change on those items that may provide useful indicators of deterioration. Therefore, what may be ‘normal’ and ‘healthy’ for one person may be considered a crisis indicator for another.

Mental health risk assessment should take into account both risk and protective factors to obtain an accurate representation of the patient as outlined in Hobb’ (1984) model (Chapter 1, section 1.0.2.3). To date the fusion of risk and protective factors for measuring crisis has not been achieved (Chapter 1, sec 1.0.9). The rating scale developed through this research offers a method for measuring risk and protective factors along the same continuum and encapsulates these considerations for every item. Clinically, this scale represents the decision considerations of mental health professionals, which provides this measure with clinical credibility.

The training programme was developed over the course of the crisis measure’s implementation across the two CRHTs. Limitations to the training programme were identified and the training programme modified to accommodate this. These changes to the training programme supported better understanding of the items and supported completion of the rating scale. This may have impacted the inter-rater reliability data for the measure due to differences in scale completion between assessors trained at different points within the research (Chapter 8, section 8.3.3). However, the crisis measure’s manual remained the same throughout the pilot and each assessor retained their own copy to refer to as needed, which may have assisted consistency in rating approach despite the changes to the training programme. It should also be recognised that the flexible scoring system utilised in this measure is a relatively new approach to scoring items and therefore orientation and confidence with this may also have impacted the inter-rater reliability.

The pilot of the crisis measure resulted in 385 completed measures and therefore surpassed our target of 300 suggested according to the rationale of Tabachnick and Fidell (2007). The next step in developing the crisis measure was to use the data obtained to understand the underlying structure of the item pool.

Chapter 3

Identifying the Key Areas of Crisis Assessment

3.1 Background

The concept of crisis assessment was comprehensively assessed in Chapter 2. The items extracted from interviews and focus groups were assessed and reduced down by the working group and a clinically credible item rating-scale was developed through information obtained in the interviews and focus groups. This process resulted in a 143 item crisis measure that utilised an 11 point *cause for concern* item rating-scale. A total of 385 crisis measures, completed by 43 members of the multidisciplinary CRHT teams were collected and the data entered into PASW/SPSS (PASW statistics 18 software, 2007 – previously SPSS) for further analysis.

The next step for this research was to assess the substance of the crisis item pool in more detail to identify the underlying structure of the crisis measure through quantitative statistical analyses as outlined in the overview of the research plan (Figure 1.8). The aim here is to identify how the items relate and identify with each other in order to assess the components of the item pool that will make up the subscales of the crisis measure (as outlined in Chapter 1, section 1.2.2). The quality of the subscale items is assessed by analysing their fit to the Rasch model. The purity of the subscales is assessed to ensure that all of the items in each subscale are significant features of a common underlying construct (unidimensionality).

It is hypothesised that the crisis measure developed through this research may be underpinned by one of three structural models (Figure 3.0). The structural model provides a manner of understanding how the items interact and provide information. The first structural model suggested (Figure 3.0.1) has one core construct to which all the items contribute information (unidimensional model). This model would provide one total score to represent an individual's level on the underlying construct. The second structural model (Figure 3.0.2) shows the items contributing information to a number of independent dimensions that would provide outcome scores for several constructs that are unrelated to each other (multidimensional model). The third model is described as a bi-factor model where items provide information to a number of subscales as well as an underpinning core dimension (DeMars, 2006). This final model would therefore provide a number of subscale scores as well as an overarching single total score.

Figure 3.0- Structural Models of Measurement

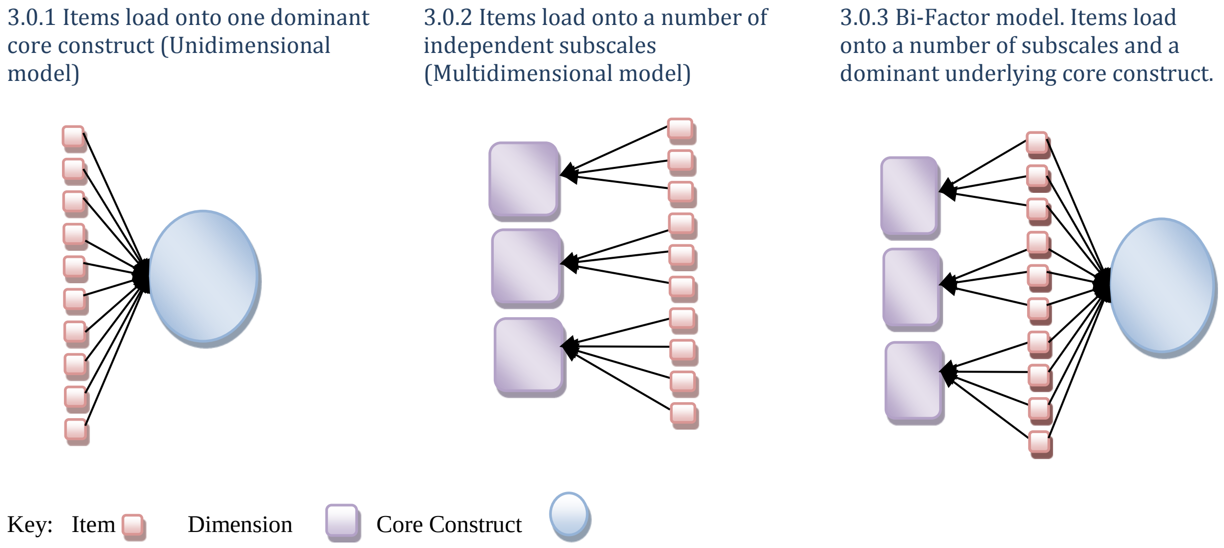


Figure 3.0: Figures outlining three possible underpinning structural models for the structure of the crisis measure developed through this research.

As previously described (Chapter 1, section 1.0.9 and Chapter 2, section 2.1) Bonynge & Thurber (2008) attempted to develop a clinical ratings scale for crisis assessment. The significant lack of construct validity outlined earlier in the methodology of this study suggests that the outcomes of their measure may not be reliable. However, their general approach of understanding the structure of the item pool using statistical analyses for identifying factors was a useful approach and provides a helpful template. Therefore, after identifying the item pool, the next step was to explore how the items relate to each other and the areas of interest they provide information for. PCA was applied to understand the structure of the crisis measure's item pool rather than Exploratory Factor Analysis as used by Bonynge & Thurber for the reasons described in Chapter 1 (1.2.2.1).

3.2 Overview of the Subscale Analyses

The aim of this section of the research was to identify the subscales of the item pool and to assess their quality. PCA is an item reduction approach that identifies the underlying component/dimension structure of an item pool. It was used in this research due to the large number of variables (items) and the belief that these variables may be reduced. The components identified from this analysis became the subscales for the final crisis measure. The scree plot produced as part of the PCA analysis provided evidence for the underlying structural model of the item pool and will be described later in this research.

Following the initial identification of the item pool subscales through PCA, goodness of fit measures used in Rasch analysis supported the choice of items and persons who demonstrated good fit to the Rasch model (as described in Chapter 1, section 1.2.1.1). Items that are shown to have good fit to the Rasch model are the items most representative within the underlying construct being measured by that group of items (Masters & Wright, 1984). Items and people shown to be problematic by how well they fit the Rasch model are treated as outliers and can be removed to reduce their influence on the outcomes of analysis, therefore enhancing the statistical qualities. This approach was used to further reduce the item pool following PCA analysis and to refine the subscales.

Finally, the identified subscales were assessed for unidimensionality through Rasch analysis. Unidimensionality, as described in Chapter 1 (sec 1.2.1.7) is important for the validity of a measure with the outcomes contributing to the overall validity evidence of this measure (see Chapter 8). Where unidimensionality is shown, it can be inferred that the items of these subscales are tapping into the same underlying construct. Once the subscales had been identified and their unidimensionality confirmed, it was the task of the researcher to identify what constructs were represented by the subscales in order to provide appropriate descriptive subscale labels. A final PCA was conducted to indicate the overall variance explained by the items remaining in the final item pool for the crisis measure. To summarise, Figure 3.1 provides an overview of the analyses.

Figure 3.1 - Flowchart of the Initial Crisis Measure Item Pool Analyses

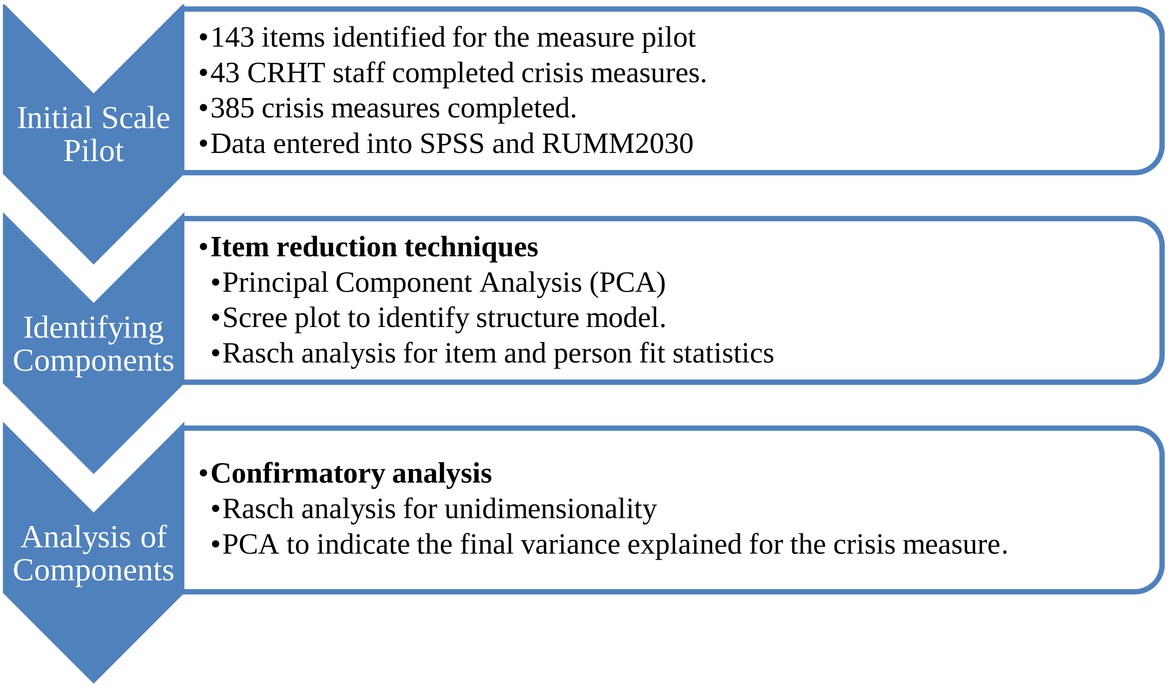
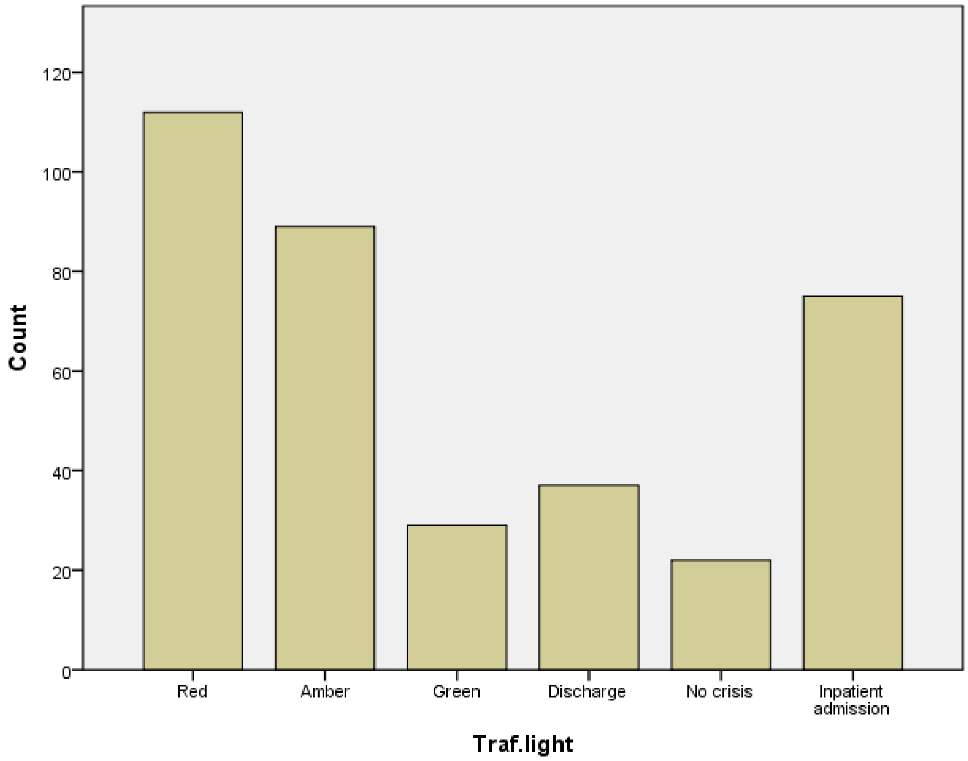


Figure 3.1 outlining the steps taken to develop the item pool for the first pilot of the crisis measure developed through this research. Three main phases of item pool development are shown here.

3.3 Data

As described in in Chapter 2 (section 2.11.4), all staff members of the Bedford and Luton CRHTs in 2007/2008 were trained to rate the item list developed from the initial interviews using the flexible *cause for concern* rating-scale developed. A total of 385 crisis measures were completed by a sample of 43 mental health workers of different professional backgrounds from the CRHT teams for Bedford and Luton. Participants rated the measure anonymously and therefore no demographic information on the sample was collected. It was agreed with the CRHT teams that information on the assessors would not be collected due to concerns around unnecessary monitoring and the use of the crisis measures' information for audit/service review when the measure was still in its pilot phase, therefore its validity and reliability were still unknown. Items were rated following initial assessment, assessment for treatment re-grading and for discharging patients. Item ratings were obtained for the full spectrum of potential crisis assessment outcomes ranging from no crisis presentation (inappropriate referrals) through to severe crisis presentation appropriate for inpatient admission (Figure 3.2). The status of the CRHT crisis assessment outcome was indicated on the completed crisis measure so that this information could be used for the purposes of assessing criterion validity at a later phase in the research (Chapter 8).

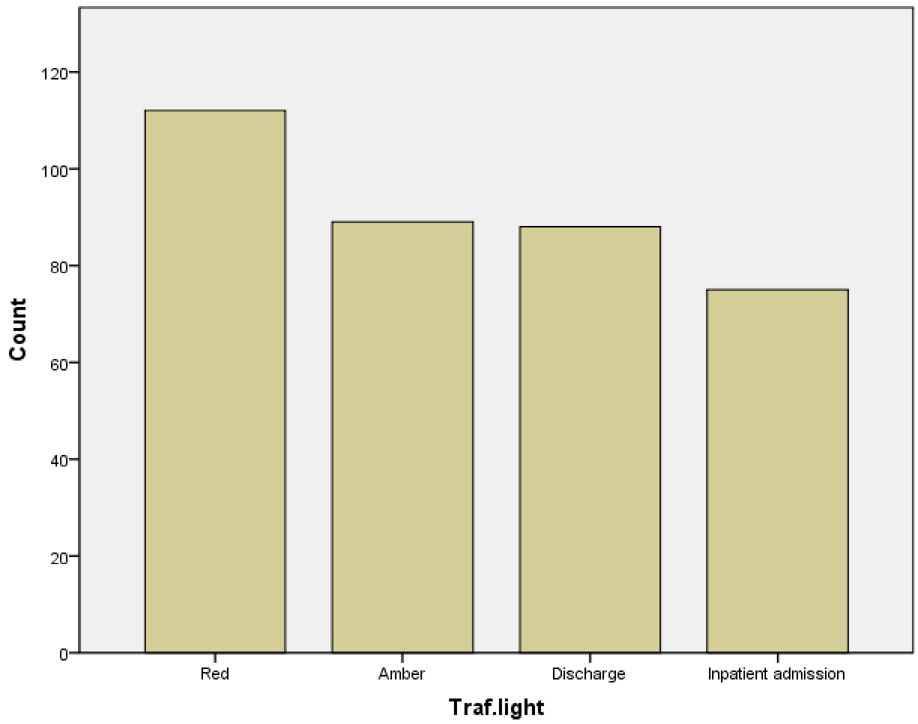
Figure 3.2 Histogram Showing the Frequency of Traffic Light Treatment Categories



Frequencies:
Red: 112 Amber: 89 Green: 28 Discharge: 38 No Crisis: 22 Inpatient Admission: 75
Missing: 21 due to treatment status not being indicated on the completed measure. Total N: 364

The number of individuals assessed as meeting the criteria for ‘Green’, ‘Discharge’ and ‘No Crisis’ treatment categories (individuals assessed as not meeting the criteria for crisis) were less than half the number assessed as meeting the criteria for Red and Amber treatment categories and therefore were not as well represented in the sample. The Green category is generally used by the CRHT teams to provide a period of monitoring to ensure that the individual is able to maintain the improvements they have shown over the course of their treatment. CRHTs do not take patients on who are assessed as 'Green' as this is viewed as a recovery or monitoring status rather than a treatment status. Therefore the crisis level indicated in the Green status is similar to those ready for discharge from the service and those assessed as not presenting in line with an acute mental health crisis presentation that requires home treatment. Based on this understanding, the categories for Green, Discharge and No Crisis were combined to represent individuals who are either ready to be or are effectively already discharged from the CRHT (Figure 3.3). This supported a more even distribution of the data across the categories which is preferable for the purposes of statistical analyses.

Figure 3.3 Histogram of the Traffic Light Treatment Categories with Collapsed Discharge Category



Frequencies:
Red: 112 Amber: 89 Discharge: 88 Inpatient-Admission: 75
Missing: 21 due to treatment status not being indicated on the completed measure. Total N: 364

3.4 Principal Component Analysis

Principal Component Analysis (PCA) was the method chosen to identify the underlying structure of the item pool due to its item reduction qualities as described in Chapter 1 (section

1.2.2.1). Through identification of the item pool components, the key areas for crisis assessment were identified and used as the subscales for the final measure.

3.4.1 Assumptions of Analysis

Data Distribution

In order to complete PCA, variables need to be continuous and normally distributed (Hatcher & Stepanski, 1994). The scale meets the criteria for continuous data distribution, however from assessing the distribution of the data it appeared that the data was skewed toward the 0 point on the item rating scales. The Shapiro-Wilks test confirmed that the variables were statistically skewed (p>0.05), which may suggest that they are not appropriate for parametric analysis. A decision needed to be made as to whether the data should be pre-processed. The skew is toward the 0 point on the scale and therefore it is likely to reduce the inter-item correlations and as a consequence downwardly bias the loadings. As a result, any pre-processing of the data to make it normally distributed would only strengthen the outcomes of PCA carried out on the data as it stands in its raw form. Therefore, where an item is shown to load onto a component in its raw form it will show stronger loadings if the data were pre-processed. In addition, research has suggested that parametric statistical approaches are robust enough to support analysis of skewed data (Gangestad & Thornhill, 1998; Glass, Peckham, & Sanders, 1972).

It was expected that the data would be skewed to the 0 point as it was unlikely that all items would be rated above 0 in the assessment of crisis patients due to the complexity and diversity of the crisis presentation. Based on this expectation, it was not surprising to find that the majority of scores from the data pool for items were on the 0 point with the remainder of the points on the scale scored approximating a normal distribution pattern. Rasch analysis does not require data to be normally distributed and therefore pre-processing was not necessary for this analysis (DeMars, 2010).

Based on the evidence regarding the strengths and robustness of parametric analysis and normally distributed data not being a requirement for Rasch analysis, it was decided to not pre-process the data and to accept it in its raw form.

Validation of the Sample Size

To complete a PCA it is suggested that a sample size of 300 is adequate (Stevens, 2002) and for Rasch analysis samples as small as 100 or 200 are often used (DeMars, 2010; Ware, Harris, Gandek, Rogers & Reese, 1997). It has been suggested that there should be at least 10 data sets per category on the item rating scale (Linacre, 2002). With an item rating scale of 11 categories this would require a minimum data set of 110 completed scales for this research. The sample collected for this analysis was 385 completed measures, which meets the assumptions for both approaches to analysis.

Outliers

PCA has been shown to be particularly sensitive to outliers (Stanimirova, Daszykowski & Walczak, 2006). Outliers are individuals who obtain a score that is very different from the majority of scores received from other individuals in the sample. Therefore, before analyses were completed histograms for each of the variables were assessed for outliers. Outliers were indicated by scores that fell at least 3 standard deviations away from the mean for the purposes of this research. Where identified the PCA was completed with and without the identified outliers to examine the impact they had on the outcome of analysis (Altman, 1991). With a sample as large as 385 and the data skewed down toward the 0 point on the majority of items it is expected that where outliers fell toward the top of the scale these would simply support to balance out the skew of the data and not overly impact the outcome of the analysis.

Outliers are excluded for the purposes of analysis using the Rasch model. Outliers were automatically identified as ‘extreme’ scores and removed by RUMM2030 software. In addition, goodness of fit to the Rasch model was assessed to remove items and persons that are shown to ‘misfit’ the Rasch model (section 3.5).

3.4.2 Initial PCA with Oblique Rotation and Item Correlation Scoping Exercise

An initial exploratory PCA was completed with the full dataset using oblique rotation. This was completed to assess if the components identified showed any relationship to each other as demonstrated through the component correlation matrix (Appendix 5). If the outcomes of this analysis demonstrated that there were no correlations between the components this would suggest that the use of orthogonal rotation would be more appropriate for the PCA analyses for this research. The PCA analysis was completed with the 143 item pool and oblique rotation was used on all components that had eigenvalues greater than 1 as indicted by the Kaiser-Guttman rule. The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, KMO = 0.933 and all KMO values for individual items were > 0.6 which is above the acceptable limit of 0.5 (Field, 2009). Bartlett's test of sphericity χ^2 (10153) = 40627.037, $p < 0.0001$, indicated that correlations between items were sufficiently large for PCA. The component correlation matrix indicated that there were no significant correlations between the 22 components identified ($r<0.5$) therefore suggesting that orthogonal rotation is appropriate for further analyses.

A data scoping exercise studied the inter-item correlation tables to identify highly correlating item pairs. Where items were shown to correlate indicates a relationship between these items. Where items correlate $r > 0.3$ this may indicate that items are repeating information already accounted for by another item, therefore potentially repeating information. This suggested that one of the items is essentially redundant (Field, 2009). Taking into account the length of the initial scale with 143 items and bearing in mind rater fatigue and bias, it was important that information was not repeated between items to ensure reliability. Where items were shown to correlate, logistic regression analysis was used to determine which item was most predictive of outcome (Field, 2009). This statistical indication of item redundancy was used in combination with information regarding the clinical credibility as well as utility and theoretical indications. This was discussed with the research team and the final decision regarding items extracted from the item pool was made on agreement. In total, 25 items were extracted as a result of this process and the remaining 118 were entered into a series of exploratory PCAs to identify the components of the item pool which make up the subscales of the crisis measure.

3.4.3 Principal Component Analysis

Orthogonal rotation was used for the initial PCA analysis as it was indicated in the interview phase of this research that both CRHT mental health professionals and patients clustered items together into separate, independent groups such as ‘harm to self’ or ‘protective factors’ and the initial oblique rotation (section 3.4.2) indicated that the components were statistically not correlated (Appendix 5).

The interview data suggested that there were unique groups of crisis indicators that could be experienced independently as well as contributing information to an overarching key variable. This suggested the presence of a bi-factor model as the underpinning structure to the item pool. Therefore the initial stage of PCA analysis was to identify these item clusters and reduce the item pool, extracting items that did not clearly load onto the components identified indicating that they were redundant from the item pool.

PCA using orthogonal rotation (varimax rotation) was applied to the data and items extracted based on their factor loadings. In basic terms factor loadings are a gauge of item importance for that component (Field, 2009). The most usual application for retaining items in a component is a factor loading of >0.3. However, the significance of a factor loading is influenced by the sample size as outlined by Stevens (2002). Based on a sample size of 300 Stevens (2002) suggested that items/variables with factor loadings greater than 0.298 should be retained.

3.4.4 Identifying the Principal Components – The Subscales.

Following rotation, items that loaded onto the component by < .390 were extracted and those loading ≥.390 were retained. The cut-off point of ≤ .390 is higher than that suggested by Stevens (2002). However due to the number of items and the generally high loading of factors, it was decided to set the cut off at a higher level to allow better discrimination between components. This pattern for loading may relate to the bi-factor model. The indication here is that items generally load onto one component more than others but demonstrate a degree of relationship to each other which may suggest that they also contribute information to an underlying core latent construct that is integral to the assessment of crisis.

Following a series of initial PCA analyses using orthogonal rotation, the item pool was reduced down to 89 items. Items statistically indicated for extraction from the item pool were discussed with the research team to consider the clinical credibility and utility of the items in addition to evidence in the literature before the final decision to extract items was made.

The PCA analysis was subsequently completed on the 89 item pool using the data sample. Following the initial PCA analysis a second analysis was conducted using varimax rotation (a method of orthogonal rotation) on all components that had eigenvalues greater than 1 as indicated by the Kaiser-Guttman rule. The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, KMO = .950 and all KMO values for individual items were > 0.6 which is above the acceptable limit of 0.5 (Field, 2009). Bartlett's test of sphericity χ^2 (3916) = 22525.483, p <0.0001, indicated that correlations between items were sufficiently large for PCA. An initial analysis was run to obtain eigenvalues for each component in the data. 13 components had eigenvalues over Kaiser's criterion of 1 and in combination explained 70.83% of the variance.

There are a number of approaches to component retention (Bartlett, 1951; Cattell, 1966; Horn 1965; Kaiser, 1960; Turner, 1998; Velicer,1976) as outlined in Chapter 1 (1.3.2.3). For the purposes of this research it was decided to use Parallel Analysis as this has been evidenced as the most accurate and therefore most representative of the underlying structure (Henson & Roberts, 2006;Horn, 1965; Turner, 1998). PCA identified a total of 13 components with eigenvalues greater than 1. The Monte-Carlo parallel analysis (Watkins, 2008) based on 100 repetitions suggested that the first 8 components should be retained for further analysis. The first 7 components had eigenvalues that fell above the cut-off for parallel analysis and therefore were automatically retained. The 8th component had an eigenvalue that fell within the parameters of the parallel analysis standard deviation and on examination of the items within this component it was agreed by the research team in consultation with the CRHTs that the 8th component should also be retained (Table 3.0). The 89 items obtained from this analysis were carried forward to the next stage of analysis - Rasch analysis for goodness of fit and unidimensionality.

Table 3.0- Outcomes for the 89 Item Principal Component Analysis

Component	PCA Eigenvalue	Parallel Analysis Eigenvalue	Parallel Analysis Std. dev	% of variance	Cumulative %	Eigenvalue after rotation	% Variance explained after rotation	Cumulative % variance explained after rotation
1	36.237	2.1371	0.492	40.715	40.715	10.766	12.097	12.097
2	5.560	2.0508	0.389	6.248	46.963	10.705	12.028	24.125
3	4.237	1.9946	0.366	4.760	51.723	8.857	9.951	34.076
4	2.720	1.9426	0.267	3.056	54.779	5.650	6.348	40.425
5	2.497	1.9000	0.260	2.805	57.584	5.128	5.762	46.187
6	2.020	1.542	0.249	2.270	59.854	3.866	4.344	50.53
7	1.959	1.8138	.0223	2.201	62.056	3.810	4.281	54.812
8	1.693	1.7753	.0203	1.902	63.958	2.997	3.368	**58.180**
9	1.365	1.7403	.0208	1.534	65.492	2.696	3.029	61.209

Monte-Carlo calculations: Number of variables: 89, Number of subjects: N= 375, Number of replications: 100

Table 3.0: indicating the 8 components retained following parallel analysis of the Principal Component Analysis identified components. The 9th component was shown to not meet the criteria for retention and was therefore dropped from the final item pool structure. The 8 components retained explained 58.2% of the variance after rotation.

3.5 Initial Rasch Analysis of Item and Person Fit Residuals Analysis – Goodness of Fit

Following the initial item reduction from the PCA analyses above (section 3.4) the research data was applied to the Rasch model using RUMM2030 (2010) software to identify any misfit between the Rasch model and the data through assessing the residual difference. The residual is the difference between the Rasch model expected item outcome and the actual item or person outcome obtained. This is explained in greater detail in Chapter 1 (section 1.2.1.1).

Item Characteristic Curves (ICC) represent the Rasch model expectations against which actual item outcomes for groups of similarly estimated *cause for concern* individuals are plotted

(DeMars, 2010). The residual is the vertical distance between the observed proportion and the expected model proportion i.e. the vertical distance between the regression line/expected score and the observed score plot.

Where the item fits the model the distance between the regression line and the observed score will be small and follow a similar curvature (Figure 3.4). Where there is greater misfit between the model and the observed score the distance between the regression line and the observed score will be greater and will follow a different curvature pattern (Figure 3.5).

Figure 3.4 - Item Characteristic Curve - Example of Good Fit

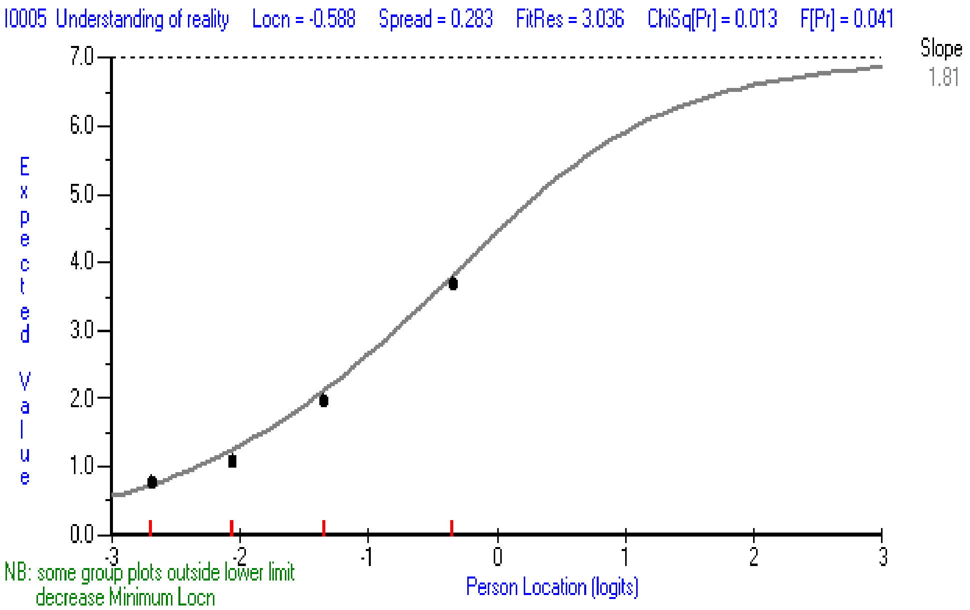


Figure 3.4 – Example of item data that fits the Rasch expected model well (regression line) as indicated by the closeness of the plots (groups) to the Rasch model s-shaped curve.

Figure 3.5 - Item Characteristic Curve - Example of Misfit

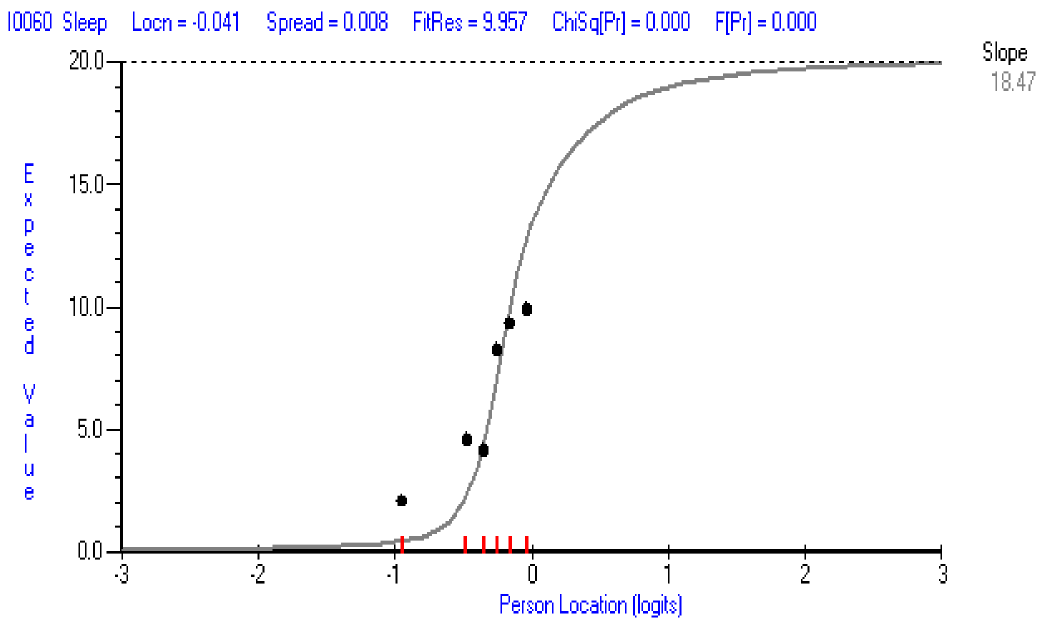


Figure 3.5 - Example of item data that does not fit the Rasch expected model (regression line) as indicated lack of coherence between the plots and the rasch s-shaped curve.

Alternatively, the individual category residual can be calculated separately (e.g. individual plots for each of the 11 categories). Although this provides much more detailed information, the graph would be complex and difficult to interpret especially for a large item measure. In addition, it has been suggested that such detailed assessment of individual categories is unnecessary as the impact on the overall component/scale would be minimal (DeMars, 2010). For this study, the residual was calculated for the predicted mean score (also called the item response function) and the observed mean score for each *cause for concern* average level group. Where particular misfit was observed, the individual category residuals were studied in more detail to look at where the particular areas of misfit lay.

Fit indices were determined by calculating a residual for each individual – the difference between the actual observed score and the expected model score. The residual was then standardised by dividing it by the standard deviation. The standardised residuals were then averaged out over persons to give item fit and averaged out over items to give the person fit (Wright & Masters, 1982).

For each component, an initial Rasch analysis was run to identify persons who did not fit the Rasch model. As recommended by Tennant (1996) persons who do not fit the Rasch model were removed before the final analysis. A fit residual cut-off of +/-3.5 was used for this sample. This cut-off is slightly larger than that normally used in analysis but is acceptable for a scale comprising of a large number of items and given that this research is at the very early stages and overly stringent criteria may result in the omission of important information from the measure. The justification provides evidence for the third reporting criterion outlined by Tennant & Conaghan (2007; Chapter 1, section 1.2.1.8) that asks for justification of the fit levels chosen for items and persons. An initial assessment of the person fit residuals was completed. The RUMM2030 software automatically removes individuals that are shown to have 'extreme' scores and then the researcher removed person data demonstrating fit residuals of +/-3.5.

The fit residuals for the items were studied and items that did not fall within the +/-3.0 fit residual were removed. Items that fell just outside of the +/- 3.0 fit residual were studied and retained if deemed clinically relevant. The fit residual cut off for items is set slightly lower than the fit residual cut off for persons to ensure that the item set covers the person's *cause for concern* spectrum. After removing misfitting items, a final total of 66 items were retained (Table 3.3).

3.6 Rasch Analysis –Unidimensionality of the Identified Components

Following the identification of the underlying component structure of the item pool and the removal of the misfitting persons and items as identified through Rasch analysis (section 3.5), the subscales were assessed for their unidimensionality (as described in Chapter 1, section 1.2.1.3), a fundamental requirement of construct validity (Fisher, 1992; Rasch, 1960; Tennant, McKenna & Hagell, 2004; Wright, 1999;). Unidimensionality indicates that the items in a subscale are measuring *one* dimension, relating to *one* construct (Hays et al, 2000). This is important based on the understanding that where items in a scale are meant to measure one construct e.g. depression, but relate to a number of constructs e.g. depression and anxiety, the outcome score will be a misrepresentation of the level of depression as it will be contaminated by the information relating to anxiety. Therefore, unidimensionality is a fundamental pre-requisite of any psychometric tool development and pivotal to the judgement of already existing measures (Tennant et al, 2004).

Rasch analysis is a confirmatory approach which has been utilised here to assess the unidimensionality of the components identified through PCA. Rasch models (Rasch, 1960) have been successfully applied to the development and validation of psychometric measures. The assessment of unidimensionality allows the researcher to add or remove items that do not fit the Rasch model to improve scale unidimensionality (Smith et al, 2006). The eight identified components/subscales were individually assessed using RUMM 2030 software (2010) for Rasch analysis for this purpose.

The unidimensionality of the identified components was tested using the method described by Smith (2002). This is outlined in Chapter 1 (section 1.2.1.3). In summary, if the component is unidimensional, once the Rasch model has been accounted for, there should be no further item associations other than by chance. Item residuals should therefore produce similar estimates of the person's *cause for concern* when compared using an independent t-test. If the t-test is significant this indicates that the items are producing different *cause for concern* estimates. However, it can still be expected that significant differences will be shown in the outcomes of the t-test ≤5% due to chance. Due to variation in the outcomes of these types of statistical analyses as a result of measurement error, sometimes significant tests will be produced for slightly more of the t-test outcomes than 5% but still indicate a unidimensional measure. When this occurs, the confidence intervals (CI) can be wrapped around the outcome using a binomial test to assess whether or not the t-test outcome falls within the parameters for the C.I. (Smith, 2002; Tennant & Conaghan, 2007; Tennant & Pallant, 2006)

Smith's (2002) approach to assessing unidimensionality was applied to the research data and where more than 5% of the t-test comparisons demonstrated significant outcomes, the binomial test was applied to take into account the confidence intervals. The outcomes demonstrate that all but one of the components indicated unidimensionality from the initial t-test analysis. Component 2 was shown to have significant t-test for >5% of the comparisons and as a consequence the binomial test was applied to the data. This showed that the t-test analysis fell within the confidence intervals for the component and therefore it was accepted as a unidimensional subscale (Table 3.1).

All of the subscales demonstrated unidimensionality, which indicates that the items contained within each of the subscales are aspects of key dimensions. The internal consistency of the subscales was assessed using Cronbach's alpha, which also demonstrated strong outcomes with all of the subscales meeting the >0.7 criteria for good reliability. The Power of Analysis of Fit statistic was shown to be good to excellent for all but one of the subscales, which suggests that that the outcomes of these analyses are reliable.

Table 3.1 – Component Unidimensionality and Reliability

Component	No. of items	**N**	No. t-tests indicating significant differences (p<0.05)	Percentage of significant t-tests	Lower 95% CI Proportion Binomial Distribution of averages	Unidimensionality Y/N	Reliability Cronbach's alpha		Power of Analysis of Fit
							With extremes	No extremes	
1	14	267	12	4.49%		Y	0.903	0.890	Excellent
2	16	272	19	6.99%	0.026	Y	0.945	0.940	Excellent
3	10	299	13	4.35%		Y	0.915	0.906	Excellent
4	6	275	13	4.73%		Y	0.865	0.808	Good
5	5	295	3	1.02%		Y	0.829	0.784	Good
6	6	268	3	1.12%		Y	0.831	0.746	Good
7	4	306	1	0.33%		Y	0.816	0.752	Good
8	5	312	6	1.92%		Y	0.781	0.708	Reasonable

Table 3.1 indicating that each of the components meets the criteria for unidimensionality as indicated by the binomial. Reliability for internal stability with Cronbach's alpha indicates that the reliability for this measure is good with and without outliers (extremes).

3.7 Principal Component Analysis - Variance Explained by the 66 Item Crisis Measure

The final PCA was run using the 66 items retained for the crisis measure to obtain an indication of the variance explained for the final item pool. Following the initial PCA analysis a second analysis was conducted using varimax rotation on all components that had eigenvalues greater than 1 (according to the Kaiser-Guttman rule - Guttman, 1954; Kaiser, 1961). The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, KMO = .957 and all KMO values for individual items were > .6 which is above the acceptable limit of .5 (Field, 2009). Bartlett's test of sphericity χ^2 (2145) = 17880.764, p <0.0001, indicated that correlations between items were sufficiently large for PCA. An initial analysis was run to obtain eigenvalues for each component in the data. Focusing on the first 8 components as retained from the previous analyses, in combination these explained 67.559% of the variance (Table 3.2). This is a reassuring outcome when compared to well established mental health outcome measure such as the Hospital Anxiety and Depression Scale, which indicates 57% of variance explained (Mykletun, Stordal, & Dahl, 2001),and the General Health Questionnaire (GHQ) where the GHQ60 item scale has been shown to indicate 64% of variance explained (Goldberg & Hillier, 1979). The item loadings did change in places, however the process of reassessing item loadings and re-running the PCA analysis could potentially result in a never ending cycle whereby the item pool is eventually reduced down further and further, subsequently limiting the richness of the information provided by the measure. Having previously removed 77 items it was important to protect the information contained within the items of the scale at this point in measurement development. It was decided to cease further analyses for item reduction until further piloting of the scale had been carried out to obtain a larger sample on which to base the statistical analyses. This will be outside of the scope of this current research. In addition, the Rasch analysis for item and person fit and unidimensionality of the subscales had been completed by this point and therefore indicated that the components were functioning well in terms of their unidimensionality.

Table 3.2- Eigenvalues and Variance Explained for the 66 Item Crisis Measure.

Component	PCA Eigenvalue	Eigenvalue after rotation	% Variance explained after rotation	Cumulative % variance explained after rotation
1	28.334	10.606	16.069	16.069
2	4.698	8.989	13.620	29.689
3	3.275	7.094	10.748	40.438
4	2.184	4.880	7.394	47.831
5	2.088	3.889	5.893	53.724
6	1.516	3.514	5.324	59.048
7	1.342	2.983	4.520	63.568
8	1.316	2.634	3.991	**67.559**

Table 3.2 indicating that the variance explained after rotation accounts for 67.559% of the variance.

3.8 Internal Consistency

Reliability is the ‘ability of a measure to produce consistent results when the same entities/constructs are measured under different conditions.’ (Field, 2009). The method adopted here was split-half reliability using Cronbach’s alpha coefficient, applied to the individual unidimensional subscales as suggested by Cronbach (1951). All subscales met an acceptable level of reliability based on Kline’s (1999) and suggested cut-offs for Cronbach’s alpha (see Table 3.1). Reliability for the whole scale (66 items) was run using PASW (formally SPSS software). Reliability for the scale was shown to be good with a Cronbach’s alpha of 0.979 (N=310).

3.9 Subscales - Titles and Composition

The final substance and structure of the item pool has been identified. The 8 components identified through these analyses will be the subscales of the resulting crisis measure. The outcomes of the PCA and Rasch analysis were discussed by the primary research team, consisting of both clinical and academic expertise, to decide on suitable subscale labels. Labels were attached to each subscale to summarise the information presented by the items (Table 3.3).

Table 3.3- Final Component Structure for the Crisis Measure

	Subscale 1 Crisis Recovery Indicators		Subscale 2 Adaptive Decision Making		Subscale 3 Risk of Harm to Self		Subscale 4 Mediating Factors		Subscale 5 Daily Structure		Subscale 6 Risk of Harm to Others		Subscale 7 Feelings/Affect		Subscale 8 Basic Needs
1	Overall thought content and clarity	17	Thought block	31	Impulsivity	41	Social Circumstances	47	Physical Exercise	52	Anger/agitation	58	Overall feelings	62	Overall appearance
2	Concentration	18	Stream of thought	32	Regret of actions during crisis	42	Protective factors	48	Isolation	53	Violence/hostility /aggression	59	Tearfulness	63	General wellbeing
3	Feelings of ineffectuality	19	Flight of ideas	33	Overall risk	43	Resourcefulness	49	Daily routine	54	Risk of neglect of others	60	Hopelessness	64	Sleep
4	Level of need	20	Poverty of thought	34	Access to lethal means	44	Daily contact with others	50	Leisure Activities	55	Family history of suicide	61	Low mood/depression	65	Appetite
5	Ability to manage symptoms	21	Understanding of reality	35	Intent to commit suicide	45	Relationships	51	Interest/ Enthusiasm	56	Risk of harm to others			66	Appropriateness of mood
6	Acceptance of difficulties	22	Capacity to consent	36	Regret of suicide attempt	46	Support Networks			57	Domestic violence				
7	Ability to relax	23	Judgement	37	Previous attempts at suicide										
8	Stability of presentation	24	Confusion	38	Risk of suicide										
9	Staff Intuition	25	Insight	39	Risk of harm to self										
10	Level of functioning	26	Irrational speech	40	Future plans										
11	Energy/get up and go	27	Overall acceptance of support												
12	Change from normal presentation	28	Ability to rationalise												
13	Predictability	29	Speech												
14	Intensity of symptoms	30	Response to hallucinations/ Delusions												
15	Responsibility for self														
16	Ability to take control														

3.10 The Subscales

The subscales identified (Table 3.3) are described in more detail below, relating the item themes to theoretical underpinnings from the literature.

3.10.1 Subscale 1 – Crisis Recovery Indicators

Thought processing, ability to cope, empowerment and stability are the main features of the recovery indicators described in subscale 1 (Table 3.3).These characteristics of the first subscale relate well to the findings of the research carried out by Tobitt & Kamboj (2011, Chapter 1, section 1.0.2.4) that investigated CRHT mental health professional's understanding of crisis. They identified the important role of functional disruption experienced by individuals in crisis in terms of the individual's ability to cope. They described the features of functional disruption as a temporary loss of ability to function, to cope, to care for oneself and to have a sense of mental control. This reflects some of the key features of this first subscale believed to represent Crisis Recovery Indicators. For example, the items of thought processing, ability to cope and stability, particularly influence a person's ability to function and to cope effectively and safely in the community setting. This directly links to the 'functional disruption' findings of Tobitt & Kamboj's study and therefore supports the validity of this study's findings. This subscale also relates well to 'The Buffering Hypothesis' (Johnson, Wood, Gooding, Taylor & Tarrier, 2011), which outlines a number of resilience factors to suicidality including attributional style, coping, problem solving, self-related beliefs and cognitive biases. This first subscale of Crisis Recovery Indicators appears to also be assessing these resilience factors or recovery indicators where coping, thought processing and empowerment appear to have a key role in determining an individual's potential for recovery from crisis.

Coping theory underpins crisis theory, describing crisis as developing from a significant break down in a person's ability to cope (Roberts & Lewis, 2001). Ability to cope was identified as a key crisis recovery indicator for mental health (Brimblecombe, 2008) and is described in subscale 1 by items that relate to the ability of a patient to manage, accept control and take responsibility for self and any difficulties they are experiencing which may help to determine how resilient a person is to crisis (items - 5,6,15,16). A person's ability to cope and manage when facing adversity or obstacles to life goals will determine how able they are to adapt and be flexible. These skills support the person to negotiate a new equilibrium as an attempt to establish a new healthy balance as outlined in the literature summarised in Chapter 1 (section 1.0.4.1). Less helpful coping strategies such as the use of disinhibitors have been shown to place individuals at heightened risk, especially in the short term (Fawcett et al, 1990) and would be represented on this measure as causing concern by the score given i.e. a score toward the *cause for concern* end of the flexible rating scale. Empowerment has been outlined as a key feature of recovery, described by Jacobson and Greenly (2001) as a method for correcting a lack of control, for moving away from a sense of helplessness and dependence towards a position of independence. Empowerment is captured in the first subscale, identifying the individual's level of need, independence and feelings of ineffectuality (items 3,4,10,15,16). Items that have concepts relating to thought processing (items - 1,2,7) link in with Subscale 2- Adaptive Decision Making - acting as a recovery indicator in the first component. As recovery progresses, symptom intensity should reduce (item 14), the crisis should dissipate and stability improve (stability – items 8, 12, 13). The score for staff intuition or 'gut-instinct' would consequently improve with improvements in the patient's presentation and other crisis recovery indicators.

3.10.2 Subscale 2 – Adaptive Decision Making

The Adaptive Decision Making subscale (items 17-30) represents the person's ability to make healthy decisions that are adaptive to functioning in the community setting. These items appear to represent three key areas: the patient's capacity to effectively process thoughts (items 17,18,19,20); the impact of thought processing on interpretation and perception (items 21,23,24,25); and, how this may be translated in terms of maladaptive behaviours (items 22,26,27,28,29,30). This links with the theme of *sense of mental control* outlined in the research of Tobitt & Kamboj (2011, Chapter 1, section 1.0.2.4). This theme has been expanded through this research and suggests that this area of assessment aims to understand a person's ability to clearly and objectively appraise situations, as influenced by their perception and interpretation, in order to make adaptive healthy decisions appropriate to living and being treated in the community setting. This reflects theory on stress and coping outlined in Chapter 1 (section 1.0.4.4) with problem solving deficits associated with exacerbating suicidal crises (Wenzel & Beck, 2008), trapping a person in a vicious cycle.

It is helpful here to consider subscale 2 in relation to decision theory. Baron (2000) outlined the 'search-inference' framework whereby the individual searches for possibilities, evidence and goals to make inferences that result in decisions. The mediating factor is judgement, described by Baron as an evaluation of possibilities in terms of presented evidence or in relation to specific goals. Judgement is influenced by an individual's ability to think rationally. Both judgement and a person's ability to be rational have been shown through research to be affected by emotion (Janis & Mann, 1977). Considering the search-inference framework in the context of crisis, it would be expected that the extreme levels of distress and emotion experienced in crisis will have a significant impact on the person's ability to be rational, therefore their ability to make sound judgement on which to base decisions. Ability to cope, which forms part of the foundation of crisis theory, would in turn influence and be influenced by the levels of distress experienced in crisis due to the impact on cognitive appraisal, ability to be rational, judgement and therefore ability to make adaptive decisions.

This subscale also relates to Freud's (1923) theory that suggests individuals only have a limited amount of psychic energy (Chapter 1, section 1.0.4.5). It could be hypothesised that an individual experiencing crisis has extremely limited levels of psychic energy which would impede their ability to understand the world in a realistic way in order to make helpful decisions and respond appropriately to situations.

3.10.3 Subscales 3 and 6 – Risk of Harm to Self and to Others

Subscale 3 (items 31-40) and subscale 6 (52-57) relate directly to the safety of the patient in terms of risk of harm to themselves and risk of harm to others which reflect the themes identified by Tobitt & Kamboj (2011, Chapter 1, section 1.0.2.4). They described an overall theme of risk of harm with a primary focus on risk of harm to self and others but with an additional interest in risk of harm from others which was highlighted from their participant interviews. The movement to home treatment as an alternative to inpatient admission for the treatment of crisis has demanded that assessment of risk is at the forefront of crisis assessment (Brimblecombe, 2008). Although positive risk may, at times, support the recovery of a patient it is also necessary to be aware of detrimental risks that may put the patient or others in the way of harm. The 2007 Mental Health Act states that compulsory admission may be sought where an individual is at risk of harm to self or others and so it is not surprising that a thorough risk assessment will form a fundamental dimension of crisis assessment for crisis teams who have the gatekeeper role for inpatient admissions. There are measures available to assess risk of suicide, for example: 'The scale of suicide ideation' (Beck, 1979) but these have not been brought together or tailored for the purposes of obtaining an overall crisis picture. This has resulted in an inconsistent approach to crisis assessment with different measures (if any measure at all) being completed for different individuals.

It was surprising to see the item 'Family History of Suicide' fall on the 'Risk of Harm to Others' subscale. However, on contemplation of the literature it was shown that risk of harm to others and risk of harm to self are often closely related and co-associated (e.g. Hillbrand, 2001) and it may be a future consideration of research with this measure to combine these subscales. This finding may link in with genetic research that has suggested individuals from families with a history of suicide may show more impulsive and aggressive behaviour linked to a genetic marker. This may consequently make them more likely to carry out self-harming behaviours (including suicide) or behaviours causing harm to others (e.g. Lopez et al, 2006).

3.10.4 Subscale 4 – Mediating Factors

Subscale 4 (41-46) describes the mediating factors that support an individual to remain at home whilst in crisis. The recovery model emphasises that considerable responsibility for recovery should be given to the individual and therefore the locus of control should be centred within the patient themselves. However, responsibility needs to be moderated in the acute phases of mental illness, where the impact of significant mental illness on thinking processes is such that the patient risks being overwhelmed by the crisis. Is it possible at this phase of illness to look toward empowerment centred in the individual or is the focus on others to support and cope for the individual as an interim measure? The acute phases of crisis may demand an initial reliance on others to cope for the patient and therefore the protective factors will significantly influence or mediate the ability for the CRHT to treat the patient at home. As the patient moves toward recovery, a shift toward the recovery model is made which empowers the patient to cope, take responsibility for managing their wellness and to function independently, reducing the emphasis on external support systems.

This links in with crisis theory and more recent research that suggests that the role of social networks and the wider community is essential for the successful resolution of crisis (e.g. Hobbs, 1984; Lindemann, 1944; Repper & Perkins, 2006; Chapter 1, section 1.0.2.2 and 1.0.2.3). The buffering hypothesis (Johnson, Wood, Gooding, Taylor & Tarrier, 2011) suggests that moderating or mediating factors should be assessed within their own right and not simply assumed by the absence of risk factors. This hypothesis concludes that the assessment and identification of moderators may help to predict suicide and therefore interventions focused on mobilising buffering factors, where these are weak, may be a powerful clinical tool. It may be that where an individual lacks the types of social supports crucial to support their recovery, the CRHT team adopt this role or act to mobilise these necessary structures for them. Where appropriate levels of community support are unavailable, inpatient admission may be the only solution. This theme was also identified by the work of Tobitt & Kamboj (2011, Chapter 1, section 1.0.2.4), who identified the theme of 'additional support needed' for the assessment of crisis. The theme of 'additional support needed' was characterised by a failure of previous support which would, as a consequence, require further support to be introduced into the individual's system to support adaptive functioning in the community setting. The Sainsbury Centre for Mental Health clearly outlines the planning and development of crisis strategies in the community setting, accessing that person's support networks to guide them towards a positive crisis resolution (Chapter 1, section 1.0.6), as the preferable approach to crisis intervention. What is clear from the outcomes of the analyses of this research in combination with previous understanding from the literature is that mediating factors such as support networks and daily contact with others are an important area for assessment that may determine the difference between an inpatient admission and home treatment.

3.10.5 Subscales 5 & 8 – Daily Structure and Basic Needs

Subscales 5 (items 47-51) and 8 (items 62-66) focus on basic levels of daily functioning. These subscales fit in with Robert's (2005) crisis definition which states that crisis results in functional impairment which acts as a significant crisis indicator. Relating this to Maslow's (1943) hierarchy of needs it would be expected that where basic needs are not satisfied, higher level needs cannot be met (Chapter 1, section 1.0.4.2). Where coping has broken down to the extent experienced by someone in acute crisis, there may be evidence of basic needs not being met and therefore therapeutic interventions are hindered until this is resolved. Again, this theme was identified in the research of Tobitt & Kamboj (2011, Chapter 1, section 1.0.2.4), who found the theme of *functional disruption* to be represented in part by an ability to care for self. Assessment of basic need and daily structure will be necessary for the development of a successful treatment intervention. This suggests that for successful crisis recovery an individual may have to be directly supported to achieve basic needs such as providing support to attend to personal hygiene and to eat a healthy balanced diet. Without these crucial building blocks, successful treatment cannot be built.

3.10.6 Subscale 7 – Feelings and Affect

Crisis theory suggests that significant threat to wellbeing and a break down in the individual's ability to cope causes feelings of significant distress (Roberts, 2005). Subscale 7 (items 58-61) looks at the feelings and affect dimension of crisis. Feelings of ineffectuality may link directly to coping theory, whereby the breakdown in the person's ability to cope results in the sense of being unable to effectively take control or problem solve. Hope is indicated as a significant feature in the recovery model (Jacobson & Greenly, 2001) and the experience of hopelessness has been indicated as a significant indicator of intent to commit suicide (Durkheim, 1952, Beck 1986) providing a powerful crisis indicator and a tool for mental health practitioners assessing crisis.

3.11 Summary and Conclusions

This research identified 8 components of crisis assessment, comprising of 66 items which accounted for 67.6% of the variance. This is statistically sound and compares favourably with symptom rating scales such as the HADS (57% of variance explained; Mykletun, Stordal, & Dahl, 2001) and the GHQ (64% of variance explained; Goldberg & Hillier, 1979) and particularly promising when compared against the previous crisis scale development project by Bonynge and Thurber (2008) that explained 43.1% of the total variance.

One of the main aims of this section in the research was to identify the underlying structure of the item pool developed in Chapter 2. Identification of the measure's structure has supported the development of subscales and indicates that the totalled subscale scores will provide a meaningful indication of the level to which a patient relates to the constructs they represent.

Reise, Morizot & Hays (2007) suggested that one of the main differences between measures was in their conceptual breadth. The conceptual breadth of a measure can be broad or narrow. For example, the measurement of depression is considered broad due to the measurement of a number of dimensions such as mood, cognitions and behaviour which are made up of a number of individual indicators. A narrower measure is one that taps into one dimension, for example a measure tapping into the construct of suicidal intent. This research illustrates that the assessment of crisis requires a broad conceptual base due to the number of subscales that were identified. This conceptual breadth will have implications for the development of a measurement tool in terms of whether Item Response Theory (IRT) or multidimensional IRT models are applied for this purpose (Reise & Waller, 2009).

Given that the conceptual breadth of crisis assessment is so broad, this shows that crisis is not a single or unidimensional construct but is comprised of a number of separate individual subscales. However, there may be a core underpinning concept or construct that is represented by these subscales when they are assessed as whole. This would be represented by a bi-factor model whereby the items in the measure provide information to a number of subscales as well as to one overarching theme. This will be explored further within the scope of this research.

The substance (items) and structure (subscales) of crisis assessment have now been identified. The next step is to refine the item-rating scales used to rate individual items to ensure that the principles of fundamental measurement are being met. In basic terms, the next step is to develop the individual item rating scales to represent true, interval level scales, with clear independent thresholds.

Following this, component cut-offs or confidence intervals can be investigated and established to complete the first steps in measurement development for a comprehensive measurement tool supporting crisis assessment.

It is important to hold in mind that this research forms the preliminary steps toward the design of a measure to encapsulate and represent the complex construct of crisis assessment. These steps are important and significant as they lay down crucial foundations upon which others may tread in making this measure more sophisticated or simpler to use. However, it is important to acknowledge the limitations of a preliminary dataset that has been based on information gathered from a small sample, collected at the very formative stages of the measure using an innovative flexible scoring approach. Implementing a measure that is in the first stages of development, using a relatively innovative approach to item rating, will not only be a challenge for the participants who are learning to implement change with a new measure, utilising a new rating method, but also the research team themselves who are developing their own understanding of the measure as it evolves from the pilot out of the initial item pool. With this in mind, however, it is encouraging that the unidimensionality of the sub-scales, the variance explained by these subscales and the reliability of this early prototype appear to indicate the valid and reliable beginnings of a very promising assessment tool for acute mental health crisis assessment.

Chapter 4

Item Refinement

Optimising the Item Rating Scales

4.1 Background

Principal Component Analysis supported a structural understanding of the item pool developed (Chapter 2) outlining 8 subscales made up of 66 items (Chapter 3). It has been helpful to assess how the items function as a whole to provide an overview of the item pool structure but it is important to assess the quality of the items at the individual level to ensure that they are providing helpful and accurate information. The aim of this chapter is to bring the focus of the research onto the individual items and their individual rating scales (level of measurement indicated in Chapter 1, Figure 1.5) in order to assess how they are functioning.

At this stage in the measurement development journey, some of the most popular measures in mental health today would have been assessed for reliability using methods embedded in CTT (Beck, Steer & Brown, 1996; Zigmond & Snaith, 1983). However, there is a fundamental difficulty in applying this traditional approach to the development of the crisis tool. This relates to a common violation of a core assumption underpinning validity and reliability. This is the violation of the assumption of interval level scaling for parametric analysis. As discussed in Chapter 1 (section 1.2.1.2) it is likely that rating scale measures developed using traditional CTT approaches will be ordinal level measures where categories on a 'scale' are ranked in order for example from low to high. It is vital that a measure developed to determine a person's level of *cause for concern* based on risk and protective factors should represent interval level scaling to ensure that validity and reliability outcomes are meaningful i.e. it measures crisis, providing an accurate picture of the crisis level and represents how the patient is coping.

To achieve interval level scaling of the global overall measure, the individual items must be functioning in a healthy manner. A healthy item scale is one that contains clear distinguishable categories (Bond & Fox, 2007; Linacre, 2006) and can provide a true representation of the item level. Therefore, to start on the path toward interval scaling, the focus at this point narrows down past the global overall measure and subscale levels to put the spotlight on the individual items and their rating scales (level of measurement as indicated in Chapter 1, Figure 1.5).

As outlined in Chapter 2 (section 2.6) the rating scale developed is one that aims to support the assessment of both risk and protective factors within the same metric. This has not been done before. Polytomous rating scales (scales with 3 or more categories) are utilised to provide richer information from items than that acquired from dichotomous response sets (Linacre, 2002) and are therefore the scale of choice for this measure. Generally the goal of implementing a rating scale is to try and capture the level or degree of an attribute and in this case the attribute is the level of *cause for concern*.

Although clinical utility is important, the item rating scale should demonstrate sound statistical qualities in terms of having clearly defined categories. The categories on the item rating scale are the units of measurement, in this case ranging from +5 to – 5. The categories on the rating scale should not overlap and should be able to provide clear and accurate representations of the person's presentation on that item (Horn, 1965). In general, reliability is shown to improve when categories are collapsed down (Stone & Wright, 1994; Zhu et al, 1997). However, this reduces the richness and depth of information provided and the scale starts to recede back to a dichotomy. The main challenge for this research is to develop a scale that is able to provide rich data in order to describe the construct in as much depth as possible whilst retaining statistical integrity. However, the statistical reality is that longer item rating scales (longer than 5 categories) result in reduced reliability. Consequently, the aim of refining the item rating scales for this measure would be to find a balance between their theoretical, clinical and statistical strengths.

As suggested in Chapter 1 (section 1.2.1.2), the equidistant steps widely used to represent rating scales do not represent the real world conceptualisation of the scale. One of the issues of contention is that the intention of measurement design may not be reflected in the reality of practice.

The intention to produce a clear unambiguous indicator of the underlying level of the construct may not necessarily be reflected in the assessors' actual conceptualisation of the scale, and more importantly the manner in which it is used, by those who utilise it. A category on a rating scale simply describes one of the steps along the scale for example 'agree' and 'disagree', and on this measure represents the levels of cause for concern ranging from +5 to -5. One category on the rating scale may represent more of the underlying construct than another category, for example the psychological leap from 'disagree' to 'agree' may be much larger than the step from 'disagree' to 'strongly disagree'. This may also occur for the crisis measure's item rating scales. Therefore the aim of this chapter is to identify how the individual item rating scales are functioning, refining item scales that are not functioning well to develop a full set of healthy, functioning and organised item rating scales which will improve the validity of the overall scale (Linacre, 2002, Wright & Masters, 1982).

Another problem that can occur is when a category is shown never to be the most probable choice along the scale continuum. This means that the category is not providing a distinct step along the continuum and therefore does not provide any further information than that provided by the other categories on the scale. When this happens, the category is said to be redundant and will be labelled as a disorganised category. In practice, identification of difficulties within the rating scale supports the identification of dysfunctional categories that can then be collapsed into 'healthier' functioning categories.

The importance of an accurate and representative rating scale cannot be overstated. As applied to measurement development and psychometric methods, the aim is to understand how legitimate it is to total scores provided by the rating scales to give an overall 'impression' of the construct being measured, to optimise the categories in order to reduce random error and therefore improve validity and reliability of the measure (e.g. Cano, Barrett, Zajicek and Hobart, 2011). As outlined in Chapter 1 (section 1.2.1), Rasch analysis (Rasch, 1960) provides a framework for assessing rating scales. Rasch analysis examines the extent to which the observed data fits with the expected data as predicted by the Rasch model. The Rasch model defines how a set of items should function if the outcomes are to be reliable and valid. Therefore, the differences indicated between the observed and Rasch expected outcomes are indicators of how rigorous the measure is.

In this section the data collected from the crisis measure pilot (Chapter 2, section 2.12) was compared to the Rasch model using a statistical programme called RUMM2030 (2010). This supported the assessment of the individual items and their rating scales. Applying the crisis measure's research data to the Rasch model will help to confirm healthy functioning item rating scales or indicate areas for improvement where scale categories are disorganised or redundant. This step is essential when developing an interval level measure.

4.2 The Choice of Rasch Model for Analysis - The Rating Scale Model Compared to the Partial Credit Model

As described in Chapter 1 (section 1.2), there are two types of polytomous model in Rasch – the rating scale model (Andrich, 1978) and the Partial Credit Model (Masters, 1989). The rating scale model uses categories/thresholds that are constant across all items and the partial credit model (PCM) assumes the categories/thresholds on individual items will differ. It was decided to apply the PCM model to the research data based on the clinical conceptualisation of individual item weightings. It is generally accepted in clinical practice that certain presenting factors/items carry more weight than others i.e. that the same score on two different items may not necessarily indicate equal levels of *cause for concern*. This concept was also indicated from the information obtained at the interview phase of this research. For example, the item 'hopelessness' was considered by clinicians to have more weight than 'low mood' based on research evidencing that 'hopelessness' is a particularly strong predictor of suicide (Beck, 1986). Due to the complexity and breadth of crisis and therefore crisis assessment it was not expected that all of the 66 items would have exactly the same 11 point scale categories. Based on the above expectations, the PCM model was adopted for analysis of the crisis measure's item data. This answers the first reporting criterion outlined by Tennant & Conaghan (2007, Chapter 1, section 1.2.1.8) that asks for justification of the Rasch model chosen for analysis.

4.3 Data

The pilot collected 385 data sets (Chapter 2) that were first entered into PASW (SPSS) and then transformed into data appropriate for RUMM2030 (2010) software. The original rating scale used had 11 categories as agreed with the staff and patient focus groups (see Chapter 2, section 2.6). For the purposes of scoring, the item rating-scale was scored 0 to 10 points with a 5 on the 'Cause for Concern' end of the scale scored as a 10 and a 5 on the 'Not cause for concern' end of the scale scored as 0. The scale uses a flexible rating system whereby the assessor can use a single score or a range covering a number of categories as outlined in Chapter 2 (2.7). When the latter approach is used (range score), the median score is taken. This could result in scores that fell between categories e.g. 4.5, 3.5 etc. RUMM2030 does not accept .5 numbers and therefore all the scores were doubled to create a 0-21 category scale in the RUMM2030 database. Following the initial rescoring and first analysis process through the RUMM2030 software, a second rescore was completed to collapse the rating scale categories back down to an 11 point scale (0-10 scale) similar to the first 11 category scoring scale. A 22 category scale (0-21) would simply be too long for use by clinicians and therefore it would not be helpful to assess the outcomes of analysis based on this presentation of the scale. Therefore, the rating scale was reduced back down to the 11 category scale following the first analysis before the results were interpreted.

4.4 Improving Item Rating Scale Categories – Category Probability Curves and Collapsing Categories

To understand how the categories for the individual item rating-scales were functioning, RUMM2030 was used to produce Category Probability Curves (CPC, Figure 4) based on the PCM model. CPCs are a plot showing the probability of each category on the item rating-scale being chosen at each step along the item's scale metric or across the variable. The metric is an interval level scale called a logit scale. As described in more detail in Chapter 1 (section 1.2.1.1), a logit scale is one where the data has been subjected to a log transformation resulting in an interval scale in which the unit intervals between points on the scale are equal (Bond & Fox, 2007). The item parameters are the positions along the measurement variable where the probability of scoring on one category is equal to the probability of scoring on the adjacent categories (Masters and Wright, 1997). This is depicted on CPCs as the intersections between the curves. In the partial credit model, the item parameters are developed by taking into account the two adjacent categories.

The category probability curves were assessed to appraise how the item rating scale categories were functioning (see Figure 4.0, for examples). Each category has its own curve on the chart representing the probability of its observation at different points along the variable. Where the curve peaks above the other category curves, this indicates that there is a greater probability of this category being observed compared to any of the other categories for that point on the variable. For a category to be considered healthy, it should be most probable at one point along the variable (Linacre, 2006; DeMars, 2010). If it is unable to achieve this, then it is not representing a separate step along the variable and does not offer any further information than that already provided in the other categories.

Guidelines suggest that there should be at least 10 observations per category, that categories should be at least 1.4 logits wide (logit as defined in Chapter 1, section 1.2.1.1) but not more than 5 logits (Linacre, 2002). However, at this very early stage of scale development, stringent application of these guidelines may result in unnecessary stripping of the crisis measure's scale, potentially losing the richness and depth of information which the scale was developed to provide in the first place. It is also important to recognise that the analysis to this point had already removed 77 items from the item pool through a number of Principal Component Analyses as well as analyses using the Rasch model. Therefore, until the 66 item crisis measure has been piloted to collect new data based on this new structure, it is safer to act on the side of caution before making any further changes to the structure or the items of the measure. In terms of construct validity, the rating scale is expected to comprehensively represent the construct of interest. The concern here is that significantly reducing the scale down at this stage without further piloting and data collection may unnecessarily reduce the scale and therefore the information it provides. At this stage of scale development, the most important criterion for each category is that it indicates a step on the variable. For a scale as complex as the *cause for concern* rating scale developed for the crisis measure (Chapter 2), this seems a reasonable cut off at this point in the measure development process.

Therefore, it was decided to implement a more basic level cut off for retaining or collapsing categories and to only ensure that the CPC for each category peaked at one point along the item rating scale continuum.

4.4.1 Item Category Analysis and Refinement

Where categories were shown to be most probable on at least one point on the variable, the category was retained. Where categories were shown to never be more probable than other categories, the category was collapsed into the adjacent category. Categories were either collapsed to support adjacent categories or where adjacent categories were shown to be strong, collapsed into the nearest/most overlapping category. For example, in Figure 4.0.1 category 5 is shown to be redundant as it does not peak above the other categories. Therefore, category 5 is collapsed down into either category 4 or 6. Category 5 was collapsed into category 4 to help enhance the 4th category on the scale resulting in the Category Probability Curves shown in 4.0.2.

Figure 4.0 - Category Probability Curves

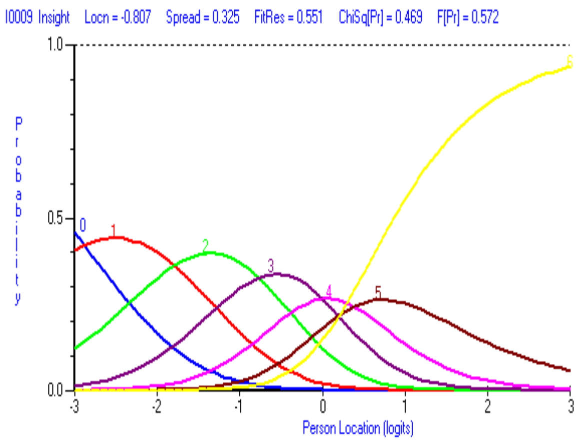


Figure 4.0.1- Example of disordered categories

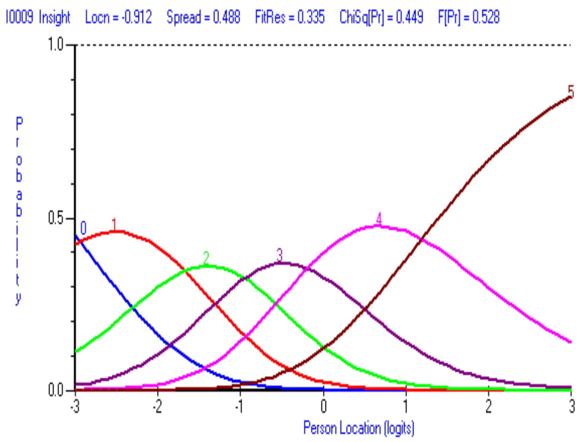


Figure 4.0.2 – Example of ordered categories

Figure 4.0: Examples of category probability curves for disordered and ordered categories. The first example (4.0.1) shows a disordered category (number 5). To correct this, the 5th category is collapsed down into the 4th category to produce the second example of category probability curves (4.0.2) which demonstrates a full set of ordered healthy functioning categories.

This process was completed with each individual item until all item categories showed at least one point of being the most probable option along the variable represented by the item.

These are summarised using threshold maps produced using RUMM2030. The threshold map outlines the threshold points between scores on the scale (see Figure 4.2 for example threshold map) The threshold is the point at which the probability of scoring a 1 for example on the scale or the next point up of a 2 becomes a ratio of 50:50. This is similar to the category probability curves but due to the mapping of all items into the same table, an overview of the transformed item scales along a single metric is captured.

The item thresholds were disordered for all items as indicated by the item threshold maps. Category Probability Curves were used to identify which categories were problematic and where to collapse unhealthy categories. Categories were collapsed by rescoring the categories in the database. An example of rescoring is shown in Figure 4.1 where category 3 was shown to be redundant and collapsed down into the second category.

Figure 4.1 – Category Rescoring

Original	0	1	2	3	4	5	6
Rescore	0	1	2	2	3	4	5

Figure 4.1. showing a typical rescore in RUMM2030 software. Here the third category is collapsed down into the second category resulting in five categories on the rating final item rating scale.

Once the rescoring had been completed, the Category Probability Curves (CPC) were re-evaluated. Where further category collapsing was required, rescoring was repeated as guided from the pattern of the CPCs. The final scoring structure for all items is shown on the item threshold maps in Appendix 6.

Following analysis and transformation of categories, all item categories were shown to be the most probable option at one point along the variable continuum as outlined in the subscale threshold maps in Appendix 6.

Figure 4.2 - Threshold Maps

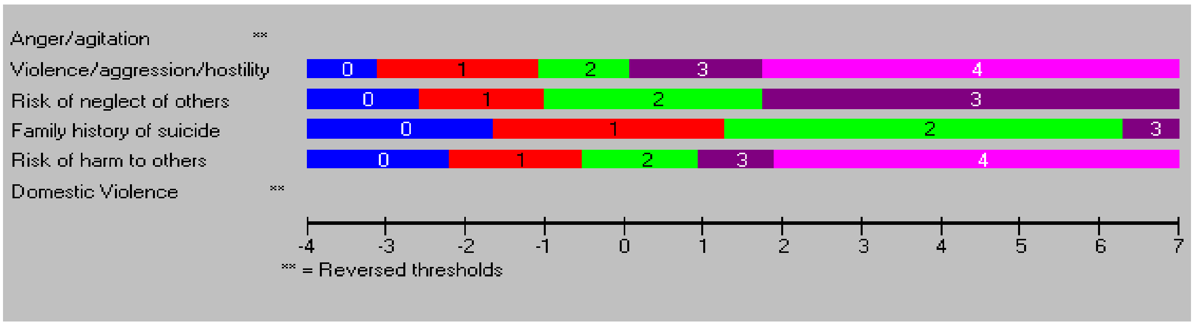


Figure 4.2.1 Example of a subscale item threshold map with disordered item rating thresholds. Two items indicate disordered thresholds as indicated by the blank line represented and two stars (**).

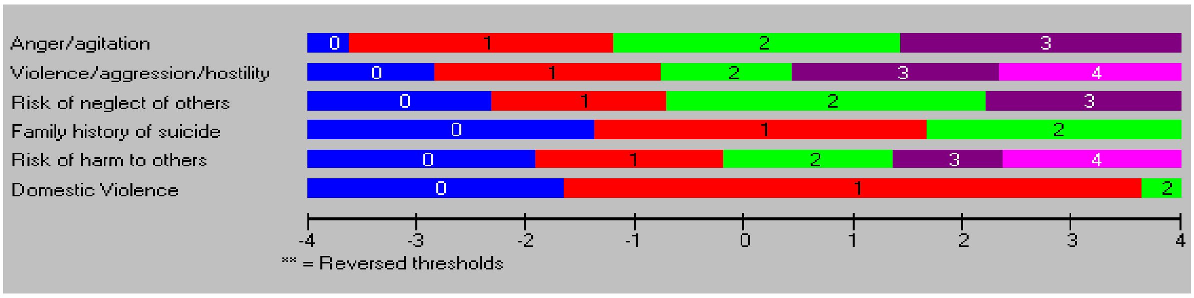


Figure 4.2.2 Example of a subscale item threshold map following the collapsing down of dysfunctional/redundant categories to produce ordered thresholds

The outcome of this analysis indicates that the 11 point *cause for concern* rating scale developed in Chapter 2 with the staff and service patient focus groups did not function in a healthy way. What this shows is that the way in which assessors perceived the scale is different to how it is depicted in the crisis measure. Therefore there is good justification for replacing the standardised item scale with individual item scales that have been tailored to each item. However, with each item having its own unique rating scale, assessors would need to continuously shift their attention between scales throughout the entirety of completing this measure. Although this may be a statistically valid approach to the problem of dysfunctional item scales, the clinical and practical validity of implementing individual item scales would suggest retaining a standardised scale across all items as it would be conceptually challenging to orientate to a different rating scale on each individual item. Therefore, it was decided to retain the original scale but to adopt a scoring system using scoring template overlays outlining the individual item scale from which staff would score the items. It would be very useful if computer software was to be developed in the future to support the scoring process.

4.5 Local Dependency

Local dependency occurs where two items tap into the same variable. Local dependency of items breaks an assumption of the Rasch model and is the fourth criterion of Tennant and Conaghan's (2007) checklist (Chapter 1, section 1.2.1.8). This assumption is that items are independent of each other and therefore should not show a relationship to each other apart from the relationship resulting from the underlying construct they represent. This assumption is outlined in Chapter 1 (section 1.2.1.4) using the example of walking, e.g. 'Can you walk 1 mile?' and asking 'Can you walk 50 meters' will show a level of dependency, i.e. if the rater indicates they can walk a mile, they will also indicate that they can walk 50 meters. "Local independence requires that the success or failure on any item should not depend on the success or failure on any other item." (Bond & Fox 2007, p172). Therefore, to ensure that the outcomes of analyses for the crisis measure are valid and meaningful, it is important to check the local dependency of items at this stage.

Item independence can be assessed by looking at the item correlations once the influence of the underlying trait has been conditioned out. If correlations are shown between the items once the underlying trait has been removed, e.g. correlations between the residuals, then this suggests that there is local dependence or may indicate that there is another underlying dimension (Lee, 2004). Based on this understanding, correlations between items should only be explained by the underlying trait of focus (Lord & Novick 1968). The main focus for the crisis measure is that variance observed is explained by the underpinning construct/s of crisis. This will ensure that the outcomes of the crisis measure support CRHT teams to make treatment decisions based on information that is directly relevant to the construct of interest, i.e. crisis. Where the assumption of local independence is violated the outcomes of analysis based on this data may be inaccurate and could indicate that the outcomes of the scale are networked into constructs other than the one of interest. At worst, a measure indicating local dependency would be providing misinformation, e.g. outcomes that could be assumed to represent the construct of crisis when in fact they are representing some other construct.

4.5.2Assessing Local Dependency

Once the underlying latent variable had been conditioned out using PCA, the item residual correlations were examined for local dependency. Item correlations were produced using RUMM2030 using Pearson's r based on the 385 completed data sets (N=385). Significant local dependency is indicated when r = +/- >0.5, moderate local dependency r = +/-> 0.4 and small local dependency r = +/- >0.3 (Siegert, Jackson, Tennant & Turner-Stokes, 2010). Some Rasch practitioners highlight correlations as small as r = +/- >0.3 for consideration but this tends to be for measures of more concrete outcomes such as those rated in physical health. For the purposes of measuring a complex construct at such an early stage of the measure's development, a correlation of r +/- > 0.3 was not considered problematic (Appendix 7).

The items in Table 4.0 were found to correlate $r > 0.3$, which may indicate a small level of local dependency however, none of the item pairs were highly correlated (r >0.50). Items that correlate $r < -0.3$ may indicate multidimensionality and are highlighted in green. However, the statistical analysis for unidimensionality has already been run (Chapter 3, section 3.6) and therefore the likeliness of multidimensionality within the individual subscales is unlikely.

Based on the understanding outlined above, none of these items were removed from the pool. No correlations > +/-0.5 were shown and therefore are within acceptable limits at this early stage of the measure development.

Table 4.0- Local Dependency Correlations

Comp	Item pair	*R*	Item pair	*r*	Item pair	*r*	Item pair	*r*
1	1. Thought block 2. Stream of thought	0.320	2. Stream of thought 3. Flight of ideas	0.399	10. Irrational speech 14. Response to hallucinations	0.350	2. Stream of thought 7. Judgement	**-0.322**
2	1. Overall thought content and clarity. 16. Ability to take control	**-0.309**						
3	3. Overall risk 6. Regret of suicide attempt	**-0.339**						
4	7. Irrational speech. 8. Response to hallucinations	0.362	1. Anger/ Agitation 2. Family history of suicide	**-0.311**				
5	1. Social circumstances 6. Support networks	**-0.333**	1. Protective factors 2. Relationships	**-0.385**				
6	1. Physical exercise 3.Daily routine	**-0.466**	1. Physical exercise 4. Leisure activities	**-0.322**	3. Daily routine 5. Interest/ enthusiasm	**-0.367**		
7	1. Overall appearance 4. Sleep	**-0.369**	1. Overall appearance 5. Appetite	**-0.322**				
8	No local dependence shown							

Table 4.0: Local dependency correlations that are greater than +/-0.3 may indicate problematic correlations between items. Where potentially problematic item correlations are shown are outlined in the table.

The correlations shown in Table 4 are all within reasonable limits and do not suggest any cause for concern for a scale attempting to measure a construct such as acute mental health crisis. The subscales all indicate unidimensionality within acceptable limits (Chapter 3, section 3.6) and the items are not demonstrating any concerning levels of dependency on each other. Based on these outcomes, the subscales can be assumed to be valid and therefore measures of reliability assumed to be accurate. Therefore no further items were removed from the item pool at this point.

4.6 Item Characteristic Curves and Item Fit Residuals

A measure is only as good as its items and the items are only as good as the rating scale which assesses them. Now that the item rating scales have been refined and local dependency assessed, the next step is to look at how the individual items are functioning overall. This is a step to ensure that the summed scale score is a good representation of the level of construct achieved. Following this, the subscale structure can be assessed to understand how well the items represent the continuum of the subscale construct.

Item Characteristics Curves (ICC) are a method of comparing the empirical item outcome against the model expected item outcome. People are grouped by similar ability (or similar level of cause for concern) and the group mean outcome is plotted for each item and compared against the Rasch expected model.

RUMM2030 produces Item Characteristic Curves (ICC) that portray how well the item data fits the Rasch model. In the example in Figure 4.3 the curved line represents the model expected outcome and the dots represent the observed scores for the person groups of different ability level. This helps to identify how discriminating the item is by examining the fit of the curve and the fit residual. Figure 4.3 is an example of where the observed outcomes compare well to the expected model outcomes. The positive Fit Residual of 0.763 indicates that the item is marginally under-discriminating.

Figure 4.3 - Item Characteristic Curve

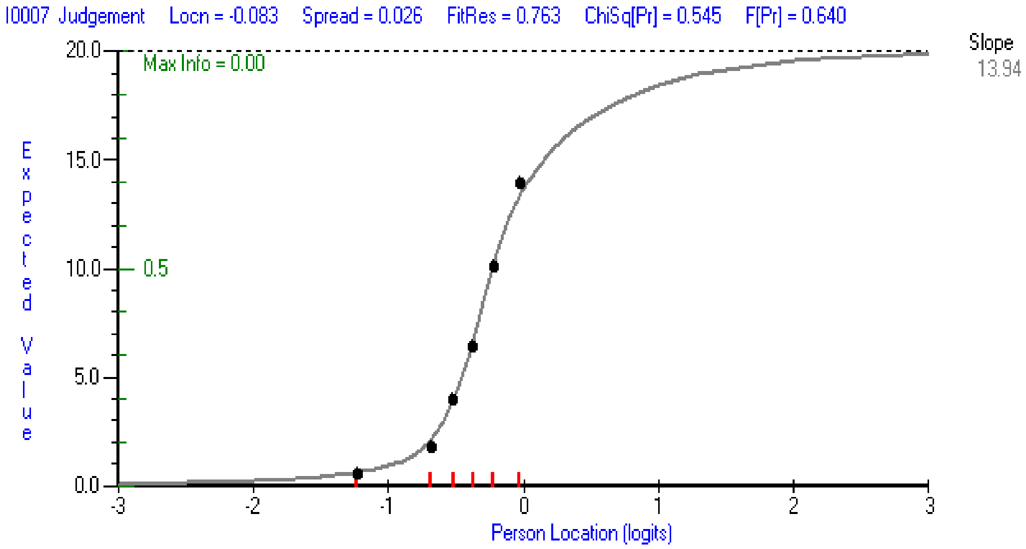


Figure 4.3: Item Characteristic Curve indicating good fit of the observed data to the Rasch expected model.

A second study of the item fit residuals was completed following the refinement of the item rating scales. This process followed a similar approach to that outlined in Chapter 3 (section 3.5). Subscales were assessed to look at the fit of items to the Rasch model using the item fit residuals. The recommended range for fit residuals is -2.50-+2.50 (Smith, 2000). However, items falling between -3.0 to +3.0 were accepted at this stage of the measure's development and because of the overall size of the scale. A second test for item fit uses the chi-square statistic which should be >0.05 if the item is to be within the acceptable parameters of the Rasch model (as described in more detail in Chapter 1, section 1.3.5.7). All fit residuals and chi-square statistics are shown in Appendix 8. Three items had fit residuals greater than +/- 3.0:

1) Understanding of reality (+3.036),

2) Capacity to consent (-3.234) and

3) Support Networks (-3.416).

Five items had fit residuals of +/-2.5: 1) Judgement (+2.810), 2). Intensity of symptoms (-2.538), 3) Previous attempts at suicide (+2.655), 4) Violence/hostility/aggression (-2.951), 5) Risk of harm to others (-2.756).

Although fit residuals for 3 of the items were shown to be above/below the ideal of +/-3.0 it was decided to retain these items as they demonstrate particular clinical relevance to the measure and their chi-square statistics were not shown to be significant indicating that these items would be able to offer useful information to the measure. It is recognised that measurement design is a balance of statistical, theoretical and clinical considerations. Ultimately, this scale is being developed to meet the need to describe the presentation of an individual experiencing crisis so that accurate treatment decisions can be made. Holding this aim in mind, it becomes clear that the items 'Understanding of reality', 'Capacity to consent' and 'Support Networks' are all important to providing a complete coherent understanding of an individual's presentation.

4.7 Distribution

Studying the distribution of the item and person locations across the same metric allows for direct comparison of how well the item set is able to represent and measure the construct of interest for the population it was designed for. Item-person distributions were mapped along the same metric for each subscale. Below is an example using subscale three, which is the subscale describing risk of harm to self. The item and the person locations are mapped against the same metric on the logit scale, which makes it easy to compare the distributions (see Figure 4.4).

Figure 4.4 - Item-Person Map

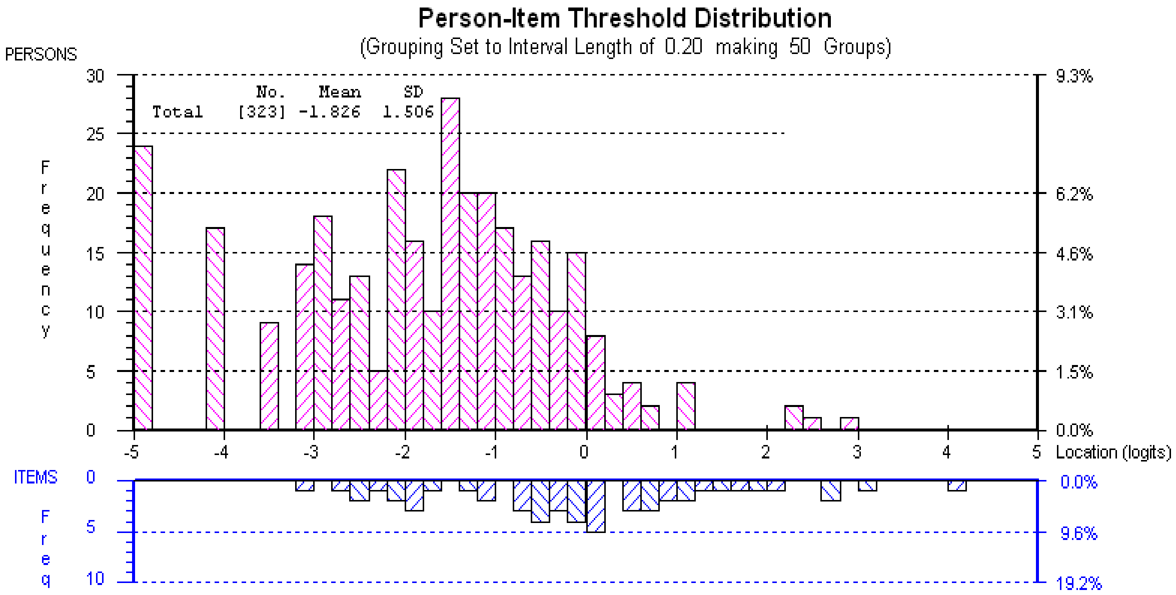


Figure 4.4: Item-Person map example. Item category locations are highlighted in blue and person locations are highlighted in red.

The statistical ideal is for the item-person distributions to mirror each other. When the item distribution is across more of the logit scale than the person distribution, items become redundant, i.e. 0% or 100% of persons achieving the item, therefore the item would not discriminate in that particular population. If the person distribution is wider than the item distribution then the items may not be able to discriminate between individuals at the outer ends of the distribution. However, there may be situations where a ceiling or floor effect is desirable and this is taken into consideration when examining the distribution. For example, the CRHT teams assess all patients referred to the service. A number of the patients referred will not require CRHT support and therefore will not score on any or very few of the items for *cause for concern* on the crisis measure. This was shown in Chapter 3, Figures 3.2 and 3.3 that showed 88 cases fell into the 'discharge' category. The final 'discharge' category includes 22 individuals who were not considered to be in crisis, 38 individuals who were ready to be discharged from the CRHT service and therefore would not be demonstrating acute or moderate crisis levels and 28 individuals who were on the Green treatment category and therefore being prepared for discharge following a period of monitoring. However, low item score outcomes are informative for a crisis measure to be used by CRHTs as they guide the mental health professional toward identifying that the patient does not require the support of the team. The same situation will also occur for patients who are at the end of their treatment with the CRHT and aiming toward discharge following a period of stabilisation. Therefore, it is expected that the person map will fall towards the lower end of the logit scale to represent the number of individuals (N=88) who are presenting with very mild or no crisis indicators. In addition, this measure has also been designed to try and discriminate between those who require inpatient assessment or admission and those who do not. This sub-sample was shown to account for 75/364 of the individuals assessed in the pilot of this measure and accounts for individuals admitted to the assessment unit for a short term assessment admission as well as individuals admitted into longer term inpatient units. The number of individuals admitted for longer term treatment will be relatively low, representing the extreme acute end of the population. The items were developed from interviews that focused on the assessment of the crisis state and therefore the items represent the levels of crisis from low to acute crisis requiring admission. It is expected that the items will fall further up the logit scale than the person population when they are mapped onto the same metric with the items spectrum representing crisis to acute crisis and the person spectrum representing no crisis to acute crisis.

Subscale Item-Person distribution maps were produced using the RUMM2030 software and evaluated to understand how well the items represent and therefore measure the person population. The items were plotted, including all of their rating scale categories, along the metric. As expected, the items were shown to be distributed across the metric but generally toward the top of the metric so are less likely to represent the lower end of the person distribution. As the tool is to be used for acute mental health crisis the items centre on capturing information on the acute and risk prevalent aspects of crisis. As a consequence the distribution found for all of the subscales represented this pattern.

In terms of producing a statistically perfect scale, the item-person maps for the subscales (Appendix 9) suggest there are difficulties around the spread of the distributions for items and persons as they are not shown to mirror each other. Statistically, when item-person distributions mirror each other, the scale is considered perfect. However, this measure is being developed for real-world practice to support the understanding of complex and often chaotic presentations which means the focus is orientated toward the clinical utility of the measure. The item-person map distributions shown in this analysis reflect what happens in the reality of clinical practice. In practice, a number of areas are assessed, not all of them relevant or concerning for each individual and this is reflected in the item-person distribution with the items starting further up the metric. In addition, not all people assessed will be appropriate for treatment under the crisis team and this is reflected in the general shift in person distribution toward the lower end of the scale.

4.8 Reliability

The focus of this chapter overall has been at the item and rating scale level. Having refined the individual items and their rating scale categories the reliability of the subscales was then assessed. This indicates how well the items are functioning together within their new rating scales. Reliability of the sub-scales is indicated in Rasch analysis by the Person Separation Indices (Table 4.1) which is the 7th criterion on the Tennant and Conaghan (2007) Rasch analysis checklist (Chapter 1, section 1.3.1.8). Similar to Cronbach's reliability, the Person Separation Indices should be >0.7 to be considered acceptable. Mok and Flynn (2002) define reliability in terms of the property of the sample being measured and the property of the scale being gauged. In general this scale offers more than adequate reliability.

Here the Person Separation Indices for the 8 components are reported. Basic Functioning, component 8, was shown to fall marginally below the cut-off of >0.7 at 0.69. However, with only 5 items in this subscale it is not surprising that better reliability has been difficult to achieve.

Table 4.1- Person Separation Indices

Comp. No.	Component Title	Person Separation Index
1	Recovery Indicators	0.92
2	Adaptive Decision Making	0.87
3	Risk of Harm to Self	0.88
4	Mediating Factors	0.82
5	Daily Structure	0.77
6	Risk of Harm to Others	0.88
7	Feelings & Affect	0.75
8	Basic Functioning	0.69

Table 4.1: Person Separation Indices outlined for the 8 subscales of the crisis measure.

Overall the subscales demonstrate good reliability. Subscale 8 indicates that there may be a weakness in how the items are functioning together. However, at this stage of the measure's development it may be premature to remove a whole subscale based on a 0.01 gap between the ideal reliability score and subscale eight's achieved outcome. By rounding this outcome up it would meet the criterion for 0.7 and therefore the decision was made to retain this component at this stage for the time being.

4.9 Summary and Conclusions

The main aim of this stage of the research was to analyse the individual items and their rating scales (measurement levels as outlined in Figure 1.5, Chapter 1). By refocusing in at this level using Rasch analysis, there was an opportunity to really understand how the items were functioning and to make improvements where difficulties were highlighted. The ultimate aim is to develop a reliable measure to accurately represent a person's level of crisis so that mental health professionals can make informed treatment decisions. If the measure is functioning well at its foundation, i.e. at the item level, then it provides a stable structure on which to build. This stable foundation was achieved by collapsing down redundant item rating scale categories, effectively removing categories that did not enhance the information provided by the scales and in doing so improved the overall health of the items.

Based on the sample size and recognition of the point at which this research is, the decision was made to not adhere stringently to all of the guidelines outlined in the literature. If stringency had been employed at this stage, the concern was that it could strip the measure of its richness unnecessarily. This process has been a delicate balance between clinical and statistical utility and the appreciation that statistical analyses can only provide statistical information, blind to the clinical implications or effects on practice. It was considered preferable to act with caution and flexibility rather than following strict criteria that may unnecessarily strip away the richness of this item pool

As indicated in section 4.4 all of the rating scales required transformation at some level. However, it was decided to retain the original rating scale format for the purposes of scale completion to ensure face validity but also to prevent rater bias caused from having to shift attention for every item and possible staff fatigue/boredom/annoyance at completing a measure that continuously changes for all 66 items. In addition, the vast majority of measures used in mental health have a standard format across all items and from a practical point of view this makes more sense. For the purposes of scoring, template overlays of the collapsed rating scales will be used to score the individual items. This scoring process will be a further training consideration and will need to be added into the measure’s manual. It is not anticipated that this process would be problematic as this approach to scoring measures has already been adopted by a number of scales for example the Millon Clinical Multiaxial Inventory-III (MCMI-III) (Millon, Millon, Davis & Grossman, 1997).

The items did not indicate any concerning levels of local dependency and they demonstrated good fit to the Rasch model following the refinement of the item scales. However, the item-person distributions did not demonstrate a statistically perfect mirror effect as suggested necessary in the literature. The item-person maps for this measure were expected to be ‘top heavy’ for the item distribution and bottom heavy for the person distribution. This was expected as the focus of the CRHT mental health professional is on the acute, more risky end of the crisis spectrum due to this sub-population requiring more intensive input and support. In reality, the CRHTs assess and treat a range of individuals that can fall at any point along the crisis spectrum including those who do not require CRHT support at all. Therefore the person distribution is expected to be skewed toward the lower end of the scale with the items distribution skewed towards the acute end of the crisis spectrum which would identify those most at risk.

Overall, the completed Rasch analysis was comprehensive and followed the outline described in Chapter 1 (section 1.3). The outcomes of the analysis were very promising and indicate that at this stage the measure provides a strong foundation upon which to build. The next step is to take the focus of analysis back out to see how the measure is functioning as a whole.

Chapter 5

Subscale Analysis

Defining Subscale Cut-offs and Item Indicators of Crisis and Risk

5.1 Introduction

Rasch analysis confirmed that the subscales identified for the crisis measure met the criteria for unidimensionality (Chapter 3, sec 3.6).This provides evidence to suggest that totalling the subscale Rasch transformed scores will provide meaningful information regarding the level of *cause for concern* presented for each subscale as suggested by the theory of fundamental measurement, in particular the Rasch model (Bond & Fox, 2007; DeMars, 2010). Information relating to an individual's levels of cause for concern in the eight different subscale areas may help define the interplay between areas of strength as well as areas of concern, which could assist care-planning. This would identify areas of greater concern, allowing them to be assigned more focused consideration as required. The development of clinically credible cut-off indictors for this measure has the potential to provide information to mental health practitioners which enables focused understanding at the subscale and whole scale level in addition to the information received at the item level. A cut-off here is defined as the point along the continuum of the totalled outcome that indicates a particular level of concern or risk. For example, an individual's totalled outcome score may indicate either low, moderate or high concern depending on which side of the predetermined cut-off point their outcome score falls.

Traditionally, cut-offs have been identified based on normative comparisons, using data collected from the population of interest and from the 'normal' population to provide a sample of expected norms. Generally a community sample is used for the 'normal' population (Kendall & Grove, 1988). Collecting normative data provides a comparison against which an individual's score can be set with the aim of guiding the mental health professional toward decisions on whether or not an individual meets 'caseness' for the symptom/presentation of interest. The difficulty with this approach is that the comparison is only as good as the samples collected, both for the clinical group (crisis group) and for the normative population. It follows that there is potential for the data to change based on the sample collected, which in turn would change the location of the cut-offs. One of the strengths of using the Rasch model for development of measures is that the item parameters and person 'ability' (in this case a person's *cause for concern*) are separated out from each other, which means that when the sample changes, the item parameters remain constant (Bond & Fox, 2007; DeMars, 2010). The item parameters in Rasch analysis are considered to be invariant, however it is recognised that this will differ by linear transformation and due to random error. Cook, Eignor and Taft (1988) suggested that item parameters could vary, however in general the consensus is that Item Response Theory i.e. Rasch analysis is more able to maintain stable item parameters when compared to approaches used in CTT due to the person and item separation.

In CTT, the item difficulty is reliant on the sample it is developed from. Item difficulty in CTT refers to the percentage of individuals in a population who answer the item 'correctly'. An item difficulty of 0.9 would therefore mean that 90% of the population/sample answered the item correctly (Christensen, Multhaup, Nordstrom & Voss, 1991). It is clear from this simple example calculation for item difficulty that outcome will be dependent on the sample used. There is an emphasis in CTT on the importance of obtaining a representative sample in order to reduce the chances of error due to discrepancy between the actual population data and the data obtained from the sample. Another advantage of the Rasch model is that the assumptions of normality and no-guessing are not required, which is more representative of what is often observed in practice. In Rasch analysis it is assumed that error is evenly distributed across the sample (Bond & Fox, 2007).

Once the sample data collected from the crisis population has been applied to the Rasch model and the scale transformed to meet the assumption of interval level scaling, it is a valid next step to indicate subscale cut-offs based on the assumption that the subscale item scores can now be totalled. Moreover, it can be expected that these cut-offs will only vary slightly between samples of the crisis population as explained above.

5.1.1 Are Cut Offs Helpful?

There are arguments for and against providing cut offs for clinical measures (Hobart, Cano, Zajicek & Thompson, 2007; Myers & Winter, 2002; Antony, 2004). Cut offs can be clinically appealing as they provide clear guidance or parameters that may help to guide standardised treatment responses and help to indicate a level on the underlying latent crisis construct. Hobart, Cano, Zajicek & Thompson (2007) summarised three main concerns relating to the provision of scale cut-offs. Two of the concerns relate to ordinal scaling. Firstly, they questioned how meaningful it is to interpret ordinal scale data on the individual level where confidence intervals may be quite large, therefore invalidating any cut offs. The second concern relates to the inability of ordinal scaling to account for unequal scale intervals (categories). The final concern is that by reducing the scale into categories reduces the spectrum of possible outcomes and therefore reduces the richness of the data. The first two concerns relating to ordinal scaling have been addressed in this research by: 1) the decision to use the Rasch model to support transformation of the item categories to healthy functioning categories and 2) to transform the subscale total to an interval level logit scale where the steps along the scale are equidistant. The final concern regarding reduction of the richness of the data is one to which there is no clear answer. The limitations and benefits in relation to crisis and CRHT teams are briefly outlined here.

One of the limitations of using measurement cut-offs for the purposes of crisis assessment is that placing emphasis on the cut-off criterion may take away from the richness and detail of the data originally gathered due to simplification of the outcomes to a high/medium/low category for example. Categorising a person’s presentation may be seen as an attempt to shoehorn the individual into a general category that may not be adequate to describe the unique experience of that person. In an age where there is a distinct movement away from categorising or labelling individuals and towards a continuum model approach (e.g. Keyes, 2002, 2005, 2007; Power, 2009) for mental health this method of categorising people seems out of context. Mental health and mental illness are no longer seen as existing in pure isolation and measurement cut-offs may be seen as a redundant exercise. In contrast, more individualist approaches to measurement outcomes tailored to the individual are being increasingly favoured such as the Reliable Change Index (Jacobson & Truax, 1991; Wise, 2004). However, more recently there has been growing recognition that both the continuum model and differentiation between the groups of mental illness and mental health may be encapsulated in the same model, such as the two continua model suggested by Westerhof & Keyes (2010). So it appears that there is room for both approaches whereby measurement outcomes can be judged against the normative data of the population and by the individual's presentation through the analysis of change scores to provide a more focused and tailored understanding.

The benefits of using cut-offs are that they can provide a useful guideline for staff. For example they could indicate when the patient has reached 'caseness' on the underlying crisis construct in question. Summarising a large amount of data can provide an overview of the presentation which may be used hand in hand with a more individualistic approach. Ideally, every person assessed by Crisis Teams would have an unlimited resource available to entirely support a tailored individual programme. However, the reality of the National Health Service is that resources are limited. The Bedford and Luton CRHT teams often hold between 25 to 40 patients with complex difficulties at any one time and therefore resources have to be targeted where they are needed most and outcome measures have a role in not only monitoring change and recovery but also in supporting the clinical team to target resources effectively.

It can be reassuring for both the professional and the patient to understand that cut-offs have been developed from evidence provided from a large sample of individuals who have experienced very similar challenges and difficulties. Cut-offs can also be useful in providing clear indicators of change and recovery which are easier to monitor, providing a summary of all the assessment information.

In addition, providing an overview of the individual crisis presentation with a valid psychometric measure may result in more effective care planning and clearer risk minimisation. This in turn frees up more of the clinicians time, which can be directed to more direct patient care.

In summary, cut-offs have a number of strengths and may guide treatment and care planning decisions when used in concert with clinical experience and ideally a varied multi-professional team. The limitations of using cut-offs appear to be most influential when cut-offs are used in isolation rather than in harmony with other approaches to understanding information obtained through assessment. Cut-offs are helpful in providing a snapshot of the presentation but it is not expected that these would completely replace other forms of assessment, but act in support of them, helping to refine and improve current approaches, making them more efficient and effective.

5.1.2 Subscale Item Indicators

Assessment of the subscales indicated that each subscale met the criteria for unidimensionality (Chapter 3 section 3.6) along with assumptions of fit (Chapter 3, section 3.5) and item dependency (Chapter 4, section 4.5). Meeting these assumptions of the Rasch model enables the identification of a number of subscale characteristics based on how the sample data fits the Rasch model. Items that are shown to fit the Rasch model closest are most representative of the underlying construct being described by that particular subscale. These items provide key characteristics of the construct and enable key crisis indicators to be identified. It is also possible to identify the items that are less likely to receive a rating indicating significant concern (Bond & Fox, 2007). When a person receives a score on an item identified by the model as less likely or probable to be scored, it may suggest that this person is presenting as a greater risk or represents a greater *cause for concern*. On this basis, items that are identified as being the least likely to receive scores to indicate concern may function as specific risk indicators which would provide the assessor with useful presentation insights.

5.2 Overview

The aims of this section are to transform the subscale totals from ordinal level data into ‘true’ interval level logit scale data using a transformation technique embedded in the Rasch model and supported by the RUMM2030 software (2010). As outlined in Chapter 1 (section 1.2.1) individual scores only become useful information once the scale itself has been validated as providing interval level data. Therefore the cut-offs for the subscales can only be developed once the subscale total has been transformed.

Once the subscale score has been transformed to an interval level scale, the second aim is to identify subscale cut offs by applying percentiles ranks that will provide parameters for very low, low, moderate, and high classifications.

The final aim is to apply the data to the Rasch model to identify the key subscale items that are a) most representative of the subscale construct and b) least likely to receive a rating i.e. least probable to be scored as a *cause for concern,* which may provide helpful item risk indicators.

5.3 Participants and Data Collection

As outlined in Chapter 2 (sec 2.12), the crisis measure developed was piloted with the CRHT teams in Bedford and Luton by 43 mental health professionals (including mental health nurses, social workers, psychology and psychiatry) who had undergone the training designed to support accurate completion of the measure. The measure was completed across the treatment spectrum, which provided data ranging from the 'normal' population (assessed as not requiring CRHT support), across the spectrum of acute mental health crisis through to patients admitted to inpatient wards (Chapter 3, Figure 3.2 and 3.3). Three hundred and eighty five measures were completed, which met the recommended sample size for Rasch analysis (DeMars, 2010; Ware, Harris, Gandek, Rogers & Reese, 1997). The data was entered into RUMM2030 (2010) software for the purposes of Rasch analysis.

5.4 Data Distribution and Descriptive Information

The descriptive data for the subscales is outlined in Table 5.0. The data for the subscales tended to be skewed toward the 0 point and the Shapiro-Wilks test revealed that the subscales were statistically skewed (p>0.05), which may suggest that they are not appropriate for parametric analysis. Normally distributed data and interval level scaling is a requirement for CTT data analysis but not for the Rasch model (Kiseliova & Kiseliovas 2004; Slinde & Linn, 1979a). Interval level scaling was achieved as part of the analysis process. The subscale totals vary between the subscales and therefore direct comparison across subscales would not be possible if used in this form (Table 5.0). For example, subscale 1 describing 'Crisis Recovery Indicators' has a total out of 57 and subscale 8 describing 'Basic Needs' is out of 16. Therefore a score of 16 on subscale 1 is very different to a score of 16 on subscale 8. To allow for comparison across the subscales the method of developing the subscale cut-offs has to be able to take account of the different subscale total scores.

Table 5.0 – Subscale Descriptive Statistics

	Subscale 1 Recovery indicators	**Subscale 2** Adaptive decision making	**Subscale 3** Risk of harm to self	**Subscale 4** Mediating factors	**Subscale 5** Daily routine	**Subscale 6** Risk of harm to others	**Subscale 7** Feelings / Affect	**Subscale 8** Basic needs
Subscale total	57	61	52	21	22	18	13	16
Mean	15.312	10.694	11.422	6.521	5.954	3.128	2.959	3.663
Std. Deviation	10.036	10.799	8.896	5.431	4.075	3.074	2.833	2.866

Table 5.0: Descriptive statistics for the 8 identified subscales of the crisis measure highlighting the differences observed in the subscale total scores.

5.5 Subscale Transformation Using Rasch Analysis - From Ordinal to Interval Level Scaling

To provide meaningful cut-offs, interval level scaling must be achieved for each of the subscales. The Rasch model (Rasch, 1960) is a modern measurement method that provides a framework for assessing rating scales (Andrich, 1978; Bond & Fox, 2007; Pallant & Tennant 2007). Rasch analysis will be used to examine the extent to which the observed data fits with the expected data as predicted by the Rasch model. This model specifies how a set of items should function if the outcomes are to be accepted as reliable and valid in terms of accurately representing the construct of interest. The differences indicated between the observed and Rasch expected outcomes are indicators of how rigorous the measure is (Cano, Barrett, Zajicek & Hobart, 2010).

Following refinement of the rating scale at the item level, analysis confirmed that the subscales all met the criteria for unidimensionality and therefore the subscale totals could be transformed into interval level logit scales (Chapter 1, sec 1.2.1.1). It is the process of refining the individual item scales, ensuring unidimensionality of the subscales and transforming the subscale total to interval level scaling, that provides evidence for the legitimacy of accepting the totalled subscale outcome score as a representation of the actual construct level present. Following this process, cut-offs were calculated with more confidence to represent the 'true' level on the underlying latent construct of the subscale.

The subscale total scores were converted to an interval level logit scale based on the application of the data to the Rasch model using RUMM2030. The conversion table (Appendix 10A) outlines the transformations from ordinal to interval level scaling and indicates the equivalent raw score to the logit score conversions. This transformed scale provides the basis upon which percentile cut-offs were calculated. A conversion graph was generated for each subscale based on the converted raw-score to logit-score table resulting in an S-shaped curve for each subscale (Appendix 11). The utility of this curve will be outlined in greater detail in subsequent sections.

5.6 Subscale Cut-off Scores

As indicated earlier, the approach for developing subscale cut-offs must take into account the differences between the subscale totals (Table 5.0). Percentiles can be used to assist this. Their use here effectively places all the subscales onto a comparable 100 point scale which allows for comparison between subscales and supports the identification of appropriate cut-offs based on the distribution of the population. The consequences of ordinal level scaling become clear when the concept of percentiles is applied for the purposes of developing meaningful cut-offs. Dividing a scale into percentiles is based on the assumption that the resulting scale has 100 equal parts. However, if the scale is not an interval level scale, this would not be possible without first carrying out the transformation described in section 5.5. Percentiles calculated based on ordinal level data would effectively be meaningless.

Dividing the subscale's total score into percentiles will guide practitioners towards how able or how concerning an individual is for functioning adaptively, in the community within reasonable risk parameters compared to other individuals in the crisis spectrum population. Subscale7 is used here to demonstrate the method adopted for calculating the subscale cut-offs using percentiles. Based on the total score range of the items in subscale 7 (feelings and affect), a conversion graph was generated to plot the raw score against the equivalent logit score. This is seen as an S shaped curve. The logit scale is sometimes referred to as the 'Location' as it indicates a person's location on the interval level logit scale based on the raw score obtained. This is shown in Figure 5.0 below where the logit interval level scale is labelled as 'Location' on the x-axis and the raw score is on the y-axis labelled 'score'.

Figure 5.0 – S-Shaped Curve

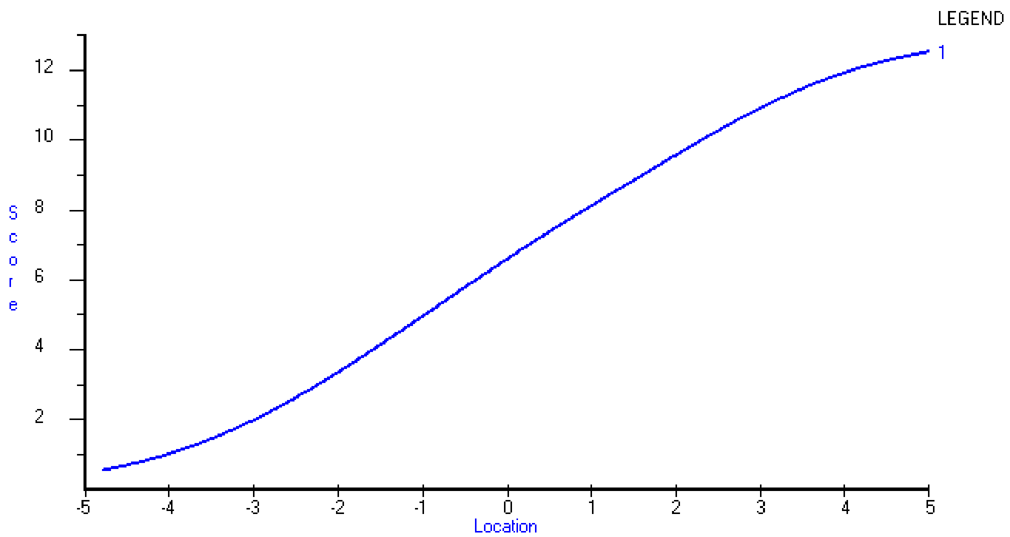


Figure 5.0: Example of an S-shaped curve. This S-shaped curve demonstrates the conversion of the raw total raw scores for subscale 7 (Feelings and Affect Subscale) mapped out against the equivalent Rasch logit scale score (interval scale). This is represented by the regression line shown as an S-shaped curve which demonstrates how the scores are unevenly distributed across the scale with scores distributed further apart at the extremes and clustered together towards the middle section of the scale.

The ‘raw score’ in this case is the raw score based on the previous transformations of the individual item scales (not the original 11 point scale used in the pilot). To calculate the true percentile bandings, the raw subscale is transformed into a logit/interval level scale (step 1, Table 5.1):

5.6.1 Step 1 – Transformation of the Raw Score to the Rasch Logit Scale

Subscale 7 has been used as the example here. The raw scale for subscale 7 is transformed into an interval level logit scale. Each point on the raw scale has an equivalent point on the logit scale as outlined in Table 5.1.

Table 5.1 – Raw Score Transformation to the Rasch Logit Scale

Subscale 7	
Raw	**Logit**
0	-4.789
1	-3.715
2	-2.860
3	-2.180
4	-1.563
5	-0.969
6	-0.384
7	0.222
8	0.893
9	1.596
10	2.263
11	2.948
12	3.783
13	4.803

Table 5.1: Example table comparing the obtained raw score on the crisis measure against the Rasch interval level logit scale. The percentile bands are highlighted:

20th < ☐ **40th <** ☐ **60th <** ☐ **80th <** ☐

5.6.2 Step 2 – Calculating Percentiles

The percentile cut-offs for Subscale 7 are calculated here as an example. The 20th percentile parameters are calculated for the subscale based on the equivalent logit scales. To calculate the percentile parameters the logit scale range is required:

- Lowest logit scale score 0 = -4.789 logits
- Highest logit scale score 13 = 4.803 logits
- Logit Scale Range: 4.803 logits - -4.789 logits = 9.592 logits

Once the logit range has been obtained, the percentile required can be calculated. The following calculation demonstrates how the 80th percentile is calculated:

- 80% cut point = lowest possible logit scale score + (logit range x 0.8)
- 80% cut point = -4.789 logits + (9.592 x 0.80) = -4.618 + 7.6736 = 3.0556 logits
- 80% cut point on the raw scale: 3.0556 logits = rounded up to raw score (r/s) of 12 = 80%

The same process is followed for the remaining percentile cut-offs which are highlighted in Table 5.2:

- 60% cut point = -4.789 + 5.7552 = 0.9662 = r/s 9
- 40% cut point = -4.789 + 3.8368 = -0.9522 = r/s 5
- 20% cut point = -4.789 + 1.9184 = -2.8706 = r/s 1

These cut-off points are based on their logit scale/true scale equivalents and will have a different outcome compared to calculating the percentile based on the raw score alone, which for the 80th percentile would be 19 x 0.80 = 15.2. This is more than a two point difference in cut-off compared to the cut-off calculated for the logit scale. It is a demonstration of how the outcomes of ordinal and interval scaling will differ, particularly significant in the context of calculating measurement cut-offs to guide mental health practitioners towards the level of *cause for concern* for an individual in crisis. The extent of this difference will depend on how closely the original scale mirrors the interval level scale. Subscale 7 is relatively small, which means that a 2 point difference in cut-off represents 10.5% of the total score. Therefore a 2 point difference in cut-off score would be significant and depicts the importance of transforming the ordinal scale to an interval scale before the scale cut-offs are applied. The S-shape curve is helpful in providing an overview of how well the raw score scale reflects the interval level logit scale. The more pronounced the S-shape curve, the less representative the raw score is of the logit scale. The less pronounced the S-shape curve, the closer the raw score scale reflects the logit scale. This can be seen in the comparison of the S-shaped curves for Subscale 2 (Adaptive Decision Making) and Subscale 3 (Risk of Harm to Self) whereby subscale 2 demonstrates a much more pronounced S-shape curve than subscale 3, which suggests that subscale 3 had a closer fit to the logit scale in its original form (as shown in Appendix 11).

The percentiles for the logit scale differ to those calculated using the raw score, which is demonstrated in the graph below (Figure 5.1). This Figure compares two different sets of 2 point raw score changes at different points along the same subscale's raw score outcome. It is clearly shown that a change in raw score by 2 points at the extremes of the raw score scale accounts for much greater change on the logit scale than a 2 point raw score increase in the middle section of the curve. Subscale 2 is used as an example below. It is shown on the graph (highlighted in red) that an increase of 2 points from a raw score of 2 to 4 results in an increase of 0.8 logits. However, when a 2 point increase in raw score is observed in the central section of the scale (highlighted in green) the logit increase is significantly less. Here an increase from 22 to 24 raw score points results in a logit increase of 0.169 logits, which is approximately an 8th of the previous increase for the same amount on the raw score scale. As the logit (location) scale is an interval level scale the difference between the two sets of score increases is comparable. This reflects the trends demonstrated in growth curves upon which the Rasch model is based.

Figure 5.1 – Subscale Raw Score Comparison to the Rasch Logit Scale Score.

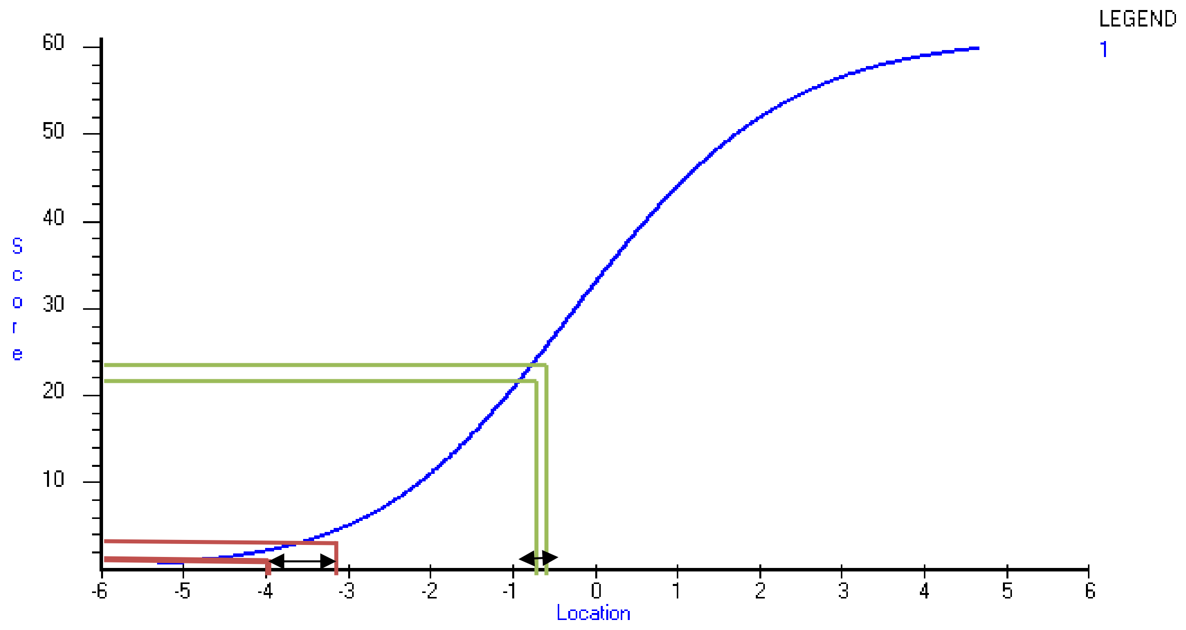


Figure 5.1: Example graph highlighting the different logit scale increase for a 2 point raw score increase at different points along the scale. X-axis is the logit scale. Y-axis is the raw score scale.

Table 5.2 below indicates the cut off points for the five percentile levels on each of the 8 subscales. The percentile cut-offs are based on interval level scaling and therefore the percentiles are comparable and will provide a good indication of individuals who indicate very low, low, moderate, high and very high cause for concern. These bandings are also highlighted in the transformation tables in Appendix 10 (Table 10A).

Table 5.2 – Subscale Percentile Cut-offs

Subscale	**Subscale 1** Recovery Indicators		**Subscale 2** Adaptive Decision Making		**Subscale 3** Risk of harm to self		**Subscale 4** Mediating Factors		**Subscale 5** Daily Structure		**Subscale 6** Risk of Harm to Others		**Subscale 7** Feeling and Affect		**Subscale 8** Basic Needs	
Percentile	logit	score <	Logit	score <	Logit	score <	logit	score <	logit	score <	logit	score <	logit	score <	logit	score <
20th	-3.254	5	-3.193	4	-2.869	5	-3.0606	3	-2.6168	4	-2.8876	2	-2.8706	1	-2.6632	2
40%	-0.406	28	-1.07	20	-0.814	18	-1.1158	8	-0.5206	10	-0.9532	6	-0.9522	5	-0.8174	6
60%	2.442	45	1.05	45	1.2404	41	0.8288	13	1.5756	16	0.9532	12	0.9662	9	1.0284	11
80%	5.29	56	3.16	58	3.295	51	2.7734	18	3.6718	20	2.8736	16	3.0556	12	2.8742	15
80th	5.30	56	≥3.17	≥58	≥3.296	≥51	2.7734	≥18	≥3.6719	≥20	≥2.8736	≥16	≥3.0556	≥12	≥2.8742	≥15

Table 5.2: Percentile Cut-offs. Table outlining the cut-offs for the 20th, 40th, 60th, 80th and 80th+ percentiles.

The percentile cut-offs calculated indicate where a person's score is compared to the rest of the distribution and so provide a good comparison of how that person is functioning in that particular subscale/construct. As the measure was piloted on a crisis population, where an individual is indicated to fall in the top two percentile bands, i.e. 80th and >80th percentile bands of 'high' and 'very high', would suggest that this is of particular concern for this person and would be a focal area for consideration in treatment planning. Subscales indicating 'very low' cause for concern would be considered areas of comparative strength for that person.

On further investigation of the percentile cut-offs it appeared that individuals achieved scores in the 80th percentile or above in less than 1% of instances and therefore the category of *very high* is unlikely to be used in clinical practice. This measure has been developed to be a pragmatic outcome tool, therefore the very high cut-off is essentially redundant, and so this was collapsed down into the cut off below it. This resulted in four remaining cut-off categories which were labelled low, moderate, high and very high.

5.7 Key Subscale Indicators

To identify the key representative items for each of the 8 subscales, the data for each subscale was applied to the Rasch model and the items assessed for how closely they fit to the Rasch model (Appendix 12). The items with the least misfit (therefore best model fit) were shown to be the most representative of the subscale's underlying construct being measured. This is assessed based on the item's fit residuals which indicate how closely items fit the model. The fit residual of an item is the difference between the Rasch model expected score and the observed score (raw score). Items that show a fit residual which is closer to 0 are demonstrating that their observed score is close to the Rasch model expected score. The closer the fit residual is to 0, the better the item fits the model, the more representative it is (DeMars, 2010). Items that were shown to have the smallest fit residuals, signalling best fit to the model for each of the 8 subscales, are summarised in Table 5.3 below and the definitions of each of the items as used for rating the crisis measure are outlined in Appendix 14.

Table 5.3 – Items Most Representative of the Crisis Subscales

Subscale No.	Subscale label	Item most representative of the subscale construct.	Fit Residual
1	Crisis Recovery Indicators	Acceptance of difficulties	-0.221
2	Adaptive Decision Making	Speech	-0.043
3	Risk of Harm to Self	Future Plans	0.174
4	Mediating Factors	Social Circumstances	0.163
5	Daily Structure	Daily routine	-0.202
6	Risk of Harm to Others	Anger and agitation	0.527
7	Mood and Affect	Low mood/depression	-0.067
8	Basic Needs	Overall Appearance	0.333

Table 5.3: Key representative subscale items. Outlining the items shown to have the smallest Fit Residuals to the Rasch model, Items with the smallest fit residuals demonstrate the closest fit to the Rasch model and can therefore be assumed as most representative of the underlying latent construct being measured.

It is interesting to note that the item *acceptance of difficulties* is shown here to be one of the strongest predictors of outcome on the first subscale, which accounts for the largest amount of variance in the crisis measure. The definition for this item on the crisis measure describes a person's ability to understand their difficulties, explaining acceptance in terms of a strength or weakness depending on the context of the difficulty being faced. Similar to other coping strategies, acceptance of difficulties can be a helpful or unhelpful approach depending on the problem being faced. The buffering hypothesis (Johnson, Wood, Gooding, Taylor & Tarrier, 2011) indicates attributional style as a significant resilience factor. Resilience describes a person's ability to positively adapt in adverse situations which may be supported by a number of helpful internal and external coping strategies. A person's attributional style describes how an individual explains their circumstances and specifically relates to how they make sense of their circumstances. The research suggests that when an individual is able to explain the events and occurrences of their crisis as external, likely to change and specific, i.e. do not generalise their specific crisis experiences, they are more likely to experience a positive resolution and to recover. In terms of the current UK treatment climate of third wave Cognitive Behavioural Therapy (CBT) approaches, there has been a shift in focus towards acceptance rather than control of difficulties. Third wave CBT approaches including Acceptance and Commitment Therapy (Hayes, Luoma, Bond, Masuda & Lillis, 2006), Dialectical Behaviour Therapy (Linehan, 1987) and Mindfulness Based Cognitive Therapy (Kabat-Zinn, 2003) have shifted the focus of therapy to bring the spotlight onto a person's relationship with their experiences by utilising therapeutic tools based on acceptance rather than resistance, avoidance or attempts at control. It seems appropriate in this context that one of the most powerful and predictive items on this measure (acceptance of difficulties) also taps into this concept of acceptance. In addition, acceptance of difficulties may also suggest recognition of the difficulties being faced and the need for change. At this time the person has moved out of their natural state of equilibrium and balance resulting in a feeling of discomfort and possibly distress. It is this feeling that opens the individual up to change (Chapter 1, section 1.0.4.1) and therefore provides a powerful indicator for crisis recovery.

The item *Speech* was indicated to be most representative of the subscale Adaptive Decision Making. The definition focuses on speech as a method for communicating and specifically the communication of needs. This may relate to either the ability to communicate either physically or verbally or the desire to communicate in terms of reaching out to others for support. In terms of treating a person in their community, it is vital that an individual is able to communicate their difficulties and needs in order to receive appropriate support. It is helpful if an individual is able to make decisions that are adaptive to the community in which the person lives as this is a particular strength and protective factor compared to when the individual is unable to effectively communicate these needs to others. Linking this back to resilience and the buffering hypothesis, social supports,and the support of family and/or a partner have been shown by a number of studies to be components of resilience that would support helpful recovery. However, a key requirement to enable support systems to be effective is the ability of the individual to communicate difficulties, to tell others when they are struggling and need additional support. The internal world of a person can only be accessed if they are able and choose to share it, and without the ability or desire to share these inner processes, an individual in crisis would be a much greater risk than an individual able to clearly communicate their needs.

The item *Future Plans* is linked to the concept of hopelessness and reasons for living (Wenzel & Beck, 2008), both of which have been described in the literature as significant indicators of risk of harm to self and in particular suicide (Baca-Garcia et al, 2004; Joiner et al, 2005, Truant et al, 1991). Future-related beliefs (Johnson, Wood, Gooding, Taylor & Tarrier, 2001) in terms of positive expectation for the future (MacLeod, Rose & Williams, 1993), optimism (Hirsch & Conner, 2006), and hope (Beck et al, 1993) have been shown to be one of the strongest predictors of suicide. Future plans as described for this measure also looks at the types of plans and goals the individual has in terms of how realistic and helpful their goals are, which may provide an indication of how helpfully the crisis will be resolved. This item will act as a key indicator of risk of harm to self for the purposes of measuring crisis using the crisis measure.

Social Circumstances such as finances, housing and work are highlighted in subscale 4 as particularly important mediating factors. Linking this back to Maslow's hierarchy of needs (1943) as outlined in Chapter 1 (section 1.0.4.2) it is clear that without the basic foundations of security in terms of shelter, food and water a person will be prevented from making any progress in their personal growth. Where an individual is experiencing particular financial difficulties, this can jeopardise their ability to access food, water and housing which ultimately acts as an obstacle to recovery, potentially maintaining the state of crisis or contributing to the development of enduring mental health problems. Where social circumstances are secure, a solid foundation is laid upon which recovery from crisis can be built. Ensuring that these basic needs are met supports motivation for recovery (Burns, Bradley & Weiner, 2012), while without the basics in place it is difficult for an individual to aspire towards higher levels and ultimately self-actualisation.

Daily routine was indicated to be the most representative item for the subscale of Daily Structure. Looking at the definition for this item (Appendix 14) it is clear that this item links in with coping theory (Chapter 1, section 1.0.4.4). The main themes of daily structure are the consequences of crisis resulting in a significant change to normal daily routines, for example sleep pattern, meal times and other general daily activities. Coping theory (Lazarus, 1993) suggests that where a person feels unable to cope and manage, they can feel overwhelmed by their difficulties and if this is not resolved, this may start to generalise to other areas of that person's life until they are unable to perform even some of the most basic routine activities. This links in with self-regulation theory (Leventhal, Brissette & Leventhal, 2003), which outlines the necessity and ability of human beings to regulate the self through behaviours that are purposeful, directed and goal oriented, generally observed through an ability to carry out basic daily tasks such as getting up in the morning, making breakfast and getting to work on time. This is achieved through basic problem solving strategies that may be compromised by significant stress events, impacting an individual's coping strategies. Where a significant change is demonstrated in an individual's ability to maintain their daily structure will signify a potential breakdown in their ability to cope and manage.

It is not surprising that the item of *anger and agitation* is the most predictive item for the subscale of Risk of Harm to Others, identified as a common sign of risk in forensic assessment (Drogin, Dattilio, Sadoff & Gutheil, 2011) and general risk assessment in mental health. It is similarly unsurprising to find that the item for *low mood/depression* is the most predictive item for the subscale of Feelings and Affect. Depression has been identified by a number of studies as a particular indicator for risk of suicide and as a consequence has regularly been identified as an exclusion criterion for research (Linehan, Comtois, Ward-Ciesielski, 2011). Depression and low mood often represent an unhappiness with the context or situation that an individual is experiencing. Where change has taken place and this change has resulted in that individual being outside of their comfort zone, low mood or depression may be observed and crisis results. This links to research on equilibrium and change that suggests that any change away from an individual's comfort zone may result in low mood or negative experience (Chapter 1, section 1.0.4.1). A person's *overall appearance* was shown to be the most representative item for the subscale of Basic Needs and may act as a summarising item that encapsulates the outcomes of an individual who has been unable to attend to their basic needs such as diet, sleep and personal hygiene.

5.8 Item Risk Indicators

Some items on the crisis measure will be less likely to receive a rating than other items, making them 'more difficult to score' when compared to other items in the measure. When these items are scored to indicate concern it is more likely that the remaining items in the subscale will also be indicated as concerning. Items that are less likely to receive a score may encompass particular risk indicators that may only apply in certain cases. If these items receive a score it may imply that the individual is particularly struggling to manage their crisis in a safe way. To identify the items that are the most difficult to score for *cause for concern* for each of the subscales, the subscale data was applied to the Rasch model and the item locations on the logit scale assessed. Items that were less probable or likely to be scored had a higher location on the logit scale compared to other items in the item pool. The greater the item location, the less likely the item is to be rated, i.e. to be rated as being a 'cause for concern'. Therefore, the item location gives an overall impression of the item difficulty (Appendix 12). In addition, the individual item rating-scale category locations were assessed using item-location maps to identify which rating-scale category was the most difficult to achieve a score (Appendix 13). For example, item X at category Y on the rating scale is least likely to be scored compared to all of the item's other categories and other item categories in the subscale, and therefore this would be considered the most 'difficult' category. This is assessing the specific point on the item's individual rating scale that is the most difficult to rate. The item identified as the most difficult item overall was not necessarily the same item containing the most difficult rating-scale category. Table 5.4 outlines the items and the item rating-scale category for each subscale that were indicated to be the least likely and therefore the most difficult to receive a score on. As before, all of the items are defined in Appendix 14.

Table 5.4 – Item Locations for Items Least Probable to Receive Ratings

Subscale No.	Subscale label	Item most difficult to receive rating overall	Item Location	Item category most difficult to rate	Item category
1	Recovery Indicators	Concentration	1.722	Concentration	Item 2 Category 3
2	Adaptive Decision Making	Speech	0.687	Ability to rationalise	Item 13 Category 5
3	Risk of Harm to Self	Future plans	0.861	Future Plans	Item 10 Category 5
4	Mediating Factors	Social Circumstances	0.499	Protective Factors	Item 2 Category 4
5	Daily Structure	Physical Exercise	0.467	Daily Routine	Item 4 Category 5
6	Risk of Harm to Others	Domestic Violence	0.906	Domestic violence	Item 6 Category 2
7	Mood/Affect	Tearfulness	0.276	Tearfulness	Item 2 Category 3
8	Basic Needs	Appetite	0.997	Appetite	Item 5 Category 4

Table 5.4: Key subscale items – table outlining the items and item rating-scale categories shown to have the highest locations on the Rasch logit scale and therefore identified as the most difficult for each subscale.

When studying the items at the rating-scale category level it can be seen that there are discrepancies between the items that present with the overall greatest ‘difficulty’ and the categories on the individual item rating scales that are the most difficult to score (as indicated in Table 5.4). For example, on subscale 2 for Adaptive Decision Making, the item *speech* is shown to be the item least likely to receive a rating of concern overall. However, it is the item *ability to rationalise’* third category (the highest point on this item’s rating scale) that is the most difficult category to score on when compared to all of the categories on all of the other items in the subscale. These may act as helpful risk indicators by highlighting to mental health team that individuals who either score on the speech item or on the highest category (category 3 on the rating scale) for the item looking at *ability* *to rationalise* would be people representing particular cause for concern. However, the reasons for items/categories being indicated as particularly 'difficult' to score on could be due to a number of possibilities and therefore should be interpreted with caution as will be discussed below.

Items that are shown to be less probable to receive a score overall and possess the most difficult rating scale category may act as a particular risk indicator for that subscale. For some of the subscales the item that is most representative of outcome is also the item that is indicated to be the least likely to receive a rating. For example, subscale 3 for 'Risk of Harm to Self' clearly identifies the item *future plans* as most representative of the subscale outcome, the item most difficult overall and also at the rating-scale category level and therefore meets all three criteria. Similarly, the item *appetite* for the subscale Basic Needs is highlighted to represent all 3 criteria. Outcomes of analysis showed that the items of *Future plans* and *Appetite* appear to be particularly prominent indicators for their subscales and therefore particular attention should be paid to these items when scored in practice.

In addition, subscales 2, 4 and 5 also indicated that the item highlighted as the most representative of the subscale outcome was also the most difficult item overall or at the rating-scale category level. Again, these items are shown to be particularly informative in terms of the subscales they represent and therefore deserve particular consideration in clinical practice.

However, it should also be recognised that items shown to be less likely or probable to receive a rating may also represent items that simply occur less often, i.e. they have a lower base rate. This presents a particular problem to this research as there is little understanding regarding the base rate occurrences of many of these items in the crisis population. Another possibility is that some of the items rated simply represent less concerning items, for example the item *physical exercise* may simply be a less concerning item for clinicians in general rather than representing a particular risk indicator when scored. Therefore it is important to assess the statistical outcomes of these findings against clinical understanding of the evidence.

Concentration was shown to be the potential risk indicator for the first subscale. On a practical level it can be appreciated that an individual who is unable to concentrate for any period of time would be at particular risk if they were to be treated in the community setting. Where significant concerns are raised regarding an individual's concentration would suggest difficulties with basic problem solving, planning and organisation which would have implications in terms of compliance with a medication regime, self-care, for example planning and making meals, safety, for example remembering to turn the cooker or taps off, attending appointments and maintaining a good daily routine to support recovery. In terms of crisis decision theory (Sweeny, 2008) an individual is expected to negotiate three steps to decision making including 1) Assess the severity of the negative event, 2) Determine response options and 3) Evaluate response options. These steps are part of the process of crisis resolution. However, an individual who is unable to concentrate is unlikely to successfully negotiate these steps without support, which would increase their risks in terms of their ability to make safe decision whilst in crisis and their ability to positively resolve the crisis in order to support future crisis resilience. Coping and problem solving have been suggested as key buffers for resilience to suicidality (Johnson et al, 2011). As suggested above, without the ability to concentrate, any attempts to problem solve or to utilise effective coping techniques would most likely be futile and therefore make the individual more vulnerable to the negative effects of crisis and risk.

Speech, future plans and social circumstances were shown to be the most representative items for the subscales that contain them (section 5.7) as well as potential key risk characteristics. The clinical understanding of these items in relation to crisis and community treatment has been outlined in the previous section and so will not be repeated here.

Interestingly, the item *physical exercise* was shown to be the item least likely to receive a rating for the subscale Daily Structure. Potentially this suggests that this item is a characteristic risk indicator. However, it may simply represent a lack of concern by clinicians for changes in exercise regime by individuals experiencing crisis when compared to other items in the item pool. From an evidence-based clinical perspective there is a growing amount of research that provides sound evidence to suggest the importance of exercise for maintaining good mental health and for preventing decline. Research to date has shown that exercise has a positive relationship with the outcome of different mental health difficulties such as depression (Blumenthal, Babyak, Doraiswamy et al, 2007), bipolar (Alsuwaidan, Kucyl, Law, & McIntyre, 2009; Sylvia, Ametrano & Nierenberg. 2010) and anxiety disorders (Strohle, 2009). Based on this evidence it could be suggested that where exercise presents itself as a cause for concern could represent a particular risk factor in terms of either a breakdown or lack of this helpful coping strategy that is likely to support a successful crisis recovery.

Domestic Violence was shown to be a characteristic risk indicator for the subscale of Risk of Harm to Others. Thinking about this item clinically it is possible that this item is simply scored less frequently because it occurs less at base rate. As described earlier, it is difficult to be certain of this as the base rate occurrences of items is unknown for the crisis population. However, it is known that domestic violence is less likely to occur in the general population. It would be interesting to know if this increases in the crisis population. In 1995 it was shown that 26% of women and 17% of men reported that they had experienced domestic violence at some point in their lives (Mirrlees-Black, 1999).

Tearfulness was shown to be the potential risk indicator for the subscale Feelings and Affect. Tearfulness is accepted as a sign of distress for a number of mental health difficulties such as adjustment disorder and depression and is a key assessment criterion for mental state examination (Oakly & Malik, 2010).

Change in *appetite* has long been associated with significant low mood and depression (Gask, Dowrick, Klinkman & Gureje, 2009) and suicidal behaviour (Cerel & Campbell, 2010), which are both significant indicators for acute mental health crisis as well as providing helpful risk indicators.

5.9 Summary and Conclusions

The rich information contained within the individual items of this construct provides the assessor with detailed pieces of a complex crisis jigsaw. It is only by bringing this information together, understanding the relationships between the pieces, that a more complete overview of the picture can be appreciated. However, the method of identifying this information so that it can later be brought together must be logical and legitimate to ensure that the final picture created makes sense and has meaning. The final interpretation of crisis will aim to bring together the information identified from the 8 subscales. What this section of the research has done is looked at how this information can be identified and summarised so that it can be brought together in a logical and legitimate way. This aim has been completed by developing cut-offs for the subscales.

Refining the individual item rating-scales (Chapter 4) provided healthy scales with clearly defined categories. Therefore the analyses completed in this section were based on the solid foundations of the refined individual item scales rather than the original 11 point *cause for concern* item rating-scales. The previous identification of unidimensional subscales provided further important evidence for the legitimacy of adding item scores together to act as indicators of the subscale construct level. The challenge for this section of the research was to transform the subscale totals from ordinal level to interval level data based on the Rasch model. Once this had been achieved it supported the implementation of percentile cut-offs to indicate the level on the construct of very low, low, moderate, high and very high. This section of the research has successfully developed informative and meaningful subscale cut-off points for this measure based on percentile calculations of the Rasch transformed interval level subscale totals.

The percentile parameters were shown to be comparable in terms of understanding areas of strength and weakness across the different subscales. Subscales indicating a low level of cause for concern (below the 40th percentile) will indicate areas of strength compared to those indicating higher levels of cause for concern (above the 60th percentile), which will be areas of particular vulnerability or weakness. Areas of vulnerability highlighted by the percentile outcomes may help to focus treatment in those particular areas, whilst areas of strength may act as protective factors that can be monitored for signs of stability or deterioration. This will be particularly useful to CRHTs in helping to direct resource in terms of the level of input required and also skill or the type of support needed, to tailor care to that particular patient.

In terms of the individual items that comprise each of the subscales, the Rasch model supported the identification of both the most representative items for each of the subscale constructs and the items that were least likely to receive a score, i.e. the most difficult to be rated for *cause for concern*. It is anticipated that identifying the items that are least likely to be scored as concerning might help future scale development by reducing the scale length. Where an individual scores high for *cause for concern* on an item least likely to receive a score of concern on a construct would make rating the remaining items on the subscale a redundant exercise as it could be predicted that the remaining items in the subscale would also receive a similarly high rating. This would significantly reduce the number of items necessary to provide information on the subscale although at the same time it should be recognised that this would further reduce the richness of the data. To use this approach would first require further research to confirm the items least likely to receive a score for each of the subscales and would be based on a much larger and therefore representative sample. This would be to ensure that decisions made on the rating of only a small subset of items would be equally as valid and reliable as rating the subscale in its entirety.

Items that have been identified as the least likely to receive ratings of significant concern may be assumed to be the most risky items, i.e. if an individual is rated as a high cause for concern on these items this may represent a significant risk for treating that person at home. However, other hypotheses may also be relevant here. For example, it may be that mental health professionals' understanding of these items prevented them from rating the item with confidence and therefore they tended to rate these items less. This links back to the importance of providing comprehensive training for staff to use this measure, especially when it is hoping to measure a presentation as complex as crisis. It could also be hypothesised that some of the items represent situations that occur at lower frequency in general, e.g. the rate of domestic violence in the general population is low and therefore less likely to be reported in general. However, currently there is little research that specifically looks at the base rate figures for the crisis population, making it impossible to ascertain whether or not an item is in fact difficult to score or has a naturally low occurrence in terms of base rate. Therefore, it is important to interpret these results with caution and flexibility. In terms of future research and future measurement development, this may provide the initial evidence from which further investigation into items that are most representative of risk can be completed .

The challenge now is to use this information to design an overall scale scoring system that will provide a summary outcome statistic. This will give an overall impression of the individual’s presentation with the aim of guiding clinical treatment decisions.

Chapter 6

Structure and Characteristics

The Crisis Risk and Adaptive Functioning Tool

6.1 Introduction

The basic structure of the item pool has been identified through PCA and the subscales have been assessed using techniques from Rasch analysis (Chapters 3-5). Specific characteristics regarding the most representative item and the least likely item to be scored as *causing concern* were identified for each subscale through the process of Rasch analysis. These analyses were possible because the subscales met the assumption of unidimensionality for Rasch analysis (Chapter 3, section 3.6), which supports meaningful comparison of item functioning. In meeting the criteria of unidimensionality for each of the subscales, evidence was provided for the legitimacy of adding the item scores together to provide a meaningful subscale total score.

The information presented by each of the subscales reflects extensive detail to understand the complexity of a person's crisis. However each subscale is unable to provide clarity concerning the overall global crisis presentation in isolation. It is only in bringing this information together, to see how this information fits, links and interweaves, that an overall understanding of the complexity of crisis can be fully appreciated. The Rasch model may offer an approach to understanding how the overall global information contained within the crisis measure relates to each other, offering a representation of the overall concept being measured. Rasch analysis would therefore act to extend the understanding of the item pool structure offered by the PCA analysis (Chapter 3),

One method for bringing the information contained within the global overall measure would be to bring this information together by simply totalling the subscale total scores to represent an overall level of crisis. The question to be answered first is how legitimate it would be to do this in order to provide an overall crisis measure total. Would such a total really be representative of a person's crisis level or overall *cause for concern*? Similar to the development of the individual subscale cut-offs, the first step is to assess how the global overall measure functions in terms of the relationships between items as a whole body, through an assessment of the dimensionality of the whole measure. Where item scores are totalled to give an impression of a construct there is an underlying assumption that all of the items inform one dimension. Therefore, the aim for this chapter is to understand the functioning of this measure as a whole in order to appreciate the qualities of the measure which may support the validity of summarising the information provided by the subscales to provide a representative overall level on the construct of interest (Thurstone, 1931b; DeMars, 2010).

This may initially appear counterintuitive based on the findings of the PCA and Rasch analysis that has shown eight separate unidimensional subscales. Brandt (2008) helpfully used the description of measuring a person's mathematical abilities to explain the concept of the bi-factor model. The aim of measurement in the context of a person's mathematical ability is to measure this single overall dimension, i.e. mathematical ability. However, mathematical ability in itself is comprised of a number of subdimensions within which an individual may demonstrate areas of strength and weakness. This describes the bi-factor model whereby an individual demonstrates an overall 'ability' or 'trait' on a general or dominant dimension of interest but also shows particular strengths or weaknesses in a number of subdimensions. For this research in particular, the overall general or dominant dimension is the level of treatment required from acute mental health services (CRHT or inpatient services) but this overall understanding is informed by a number of subdimensions across which the individual may demonstrate areas of strength or weakness and these are the eight subscales identified through the PCA (Chapter 3).

To analyse this type of model there has been the development of specific bi-factor binary item response analysis procedures (Gibbons et al, 2007). However, there has been very little development in this area for graded response data, such as the measure developed through this research, until very recently (Brandt, 2008) and these are currently only used by experts in the field of item modelling using specific software developed for this purpose.

This section of the research aimed to understand the likeliness of the item pool containing a bi-factor model structure. If a bi-factor model is indicated it would comprise of one core underpinning dimension that would represent and describe the common theme of the measurement provided by the 66 items as well as the sub-dimensions (subscales) they describe (Chapter 3). If the presence of a bi-factor model with a core dominant dimension is shown, the next step is to look at analysing the item pool as a whole in order to understand the dominant dimension's key characteristics and what it may offer in terms of expanding our knowledge about the core element of crisis assessment.

Understanding the characteristics and qualities of crisis will support an accurate appreciation of how meaningful the summed subscale outcome data will be for the purposes of guiding treatment decisions. However, this research aims to provide a clinically meaningful tool as well as a statistically accurate measure which requires the continuous critical appraisal of the outcomes of analysis to ensure that the measure is clinically relevant and meets the ultimate aim of this research, which is to support crisis assessment treatment decision making.

Therefore the aims of this section are to:

1. Assess the dimensionality of the crisis measure by analysing the overall structure utilising techniques from CTT and Rasch analysis (section 6.2).
2. Identify the core characteristics of the crisis measure, i.e. the items that are most representative of crisis assessment and the items least likely to receive a rating for *cause for concern* (most difficult items to receive a rating) (section 6.3).
3. Appraise the usefulness and clinical relevance of the crisis measure's total score for understanding a person's crisis presentation and for the purposes of comparison across time intervals or between individuals (section 6.4).

4. Describe the primary dominant dimension of the crisis measure as identified by the scree plot based on the information obtained from the above analyses (section 6.5).

6.2 Assessing the Model Structure of the Crisis Measure

To obtain a basic understanding of the crisis measure's structure there are two main techniques that will be used here; analysis of the scree plot and statistical analysis of unidimensionality using Smith's (2002) approach. There are more complex techniques embedded in Structural Equation Modelling (e.g. Pearl, 2000; Simon, 1953 ; Wright, 1921) which enable comprehensive modelling of measures such as the crisis measure. Structural Equation Modelling is a useful technique for identifying latent variables that are not directly assessed by the measure but are exposed through understanding the variables' intercorrelations (Ullman, 2006). This analysis can uncover latent psychological processes that can then be further investigated through research (Bornstein, 2011). However, this would be used in this case as a confirmatory technique to test hypotheses regarding the structure of the crisis measure following a much more substantial pilot to obtain a dataset of at least 660 based on the 66 item measure (calculated from Nunnally, 1978), although a sample closer to 1000 would provide more confidence in the findings (Comrey & Lee, 1992; Tabachnick & Fidell, 2007). A large scale pilot is outside of the scope of this research. This section will generate the structural model hypotheses upon which future Structural Equation Modelling may be used to assess these hypotheses and confirm a final structural model.

6.2.1 Analysis of the Scree Plot

One method for analysing the dimensionality of an item set is to study the eigenvalues of the inter-item correlation matrix. As with PCA, there is an assumption of normality but for the same reasons stated in Chapter 3 (section 3.3), data was not pre-processed here even though the item data spread was not normally distributed. This is due to the data being skewed to the zero point and therefore likely to reduce the inter-item correlations, downwardly biasing the loadings. Therefore pre-processing the data would only strengthen the outcomes of PCA and eigenvalues obtained for the scree plot (Figure 6.0).

Figure 6.0 - Scree Plot of the 66 Item Crisis Measure

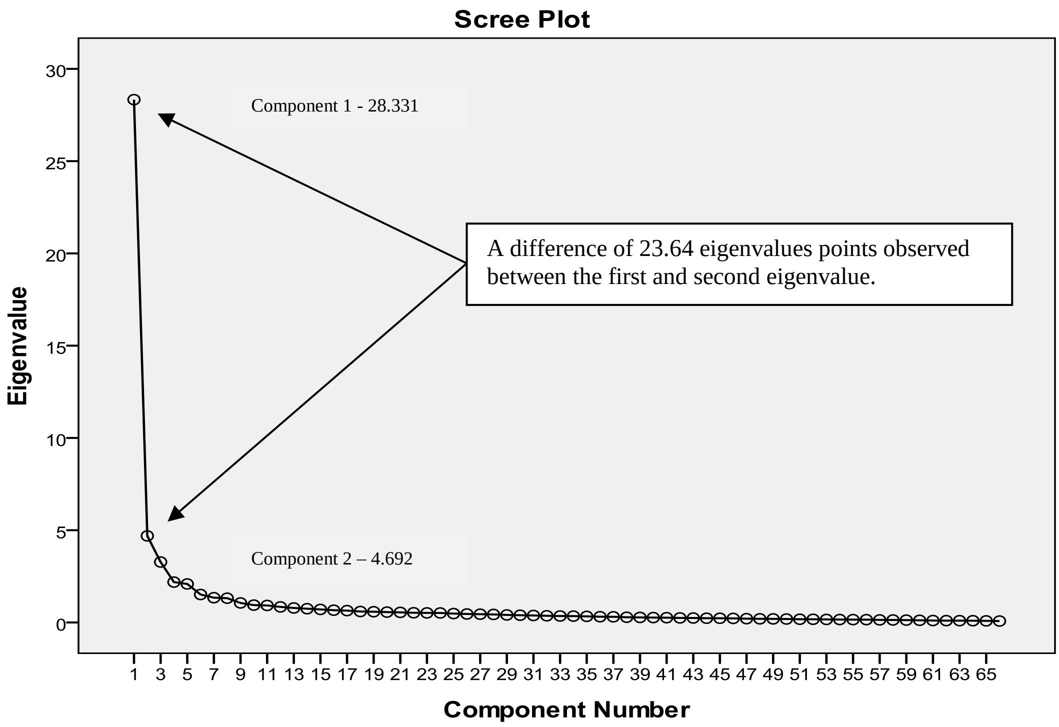


Figure 6.0 The scree plot indicates a difference of 23.64 eigenvalues between the first and second eigenvalue on the scree plot. The second component is only 16.6% of the first component.

The most common method for studying the eigenvalues is to plot the eigenvalues onto a graph called a scree plot. The general pattern of a scree plot is a steep drop at a point between eigenvalues with the rest levelling off at the bottom (the scree effect). To analyse the scree plot for the number of dimensions in the crisis measure, the number of eigenvalues before the steep drop are counted as representing the dimensionality of the of the overall item pool. Hambleton and Rovinelli (1986) commented that use of the scree plot for analysis of dimensionality was likely to indicate too many dimensions. However, due to the previous use of parallel analysis (Chapter 3, sec 3.4.4) to identify the number of dimensions to retain, the focus here will simply be on the pattern of the eigenvalues. Zwick &Velicer (1986) described the scree plot test of dimensionality as “generally accurate but variable". It has been suggested that all factors/dimensions with eigenvalues greater than one should be retained but this has been found to retain extra factors (Reckase, 1979) and therefore the more direct approach of studying the scree plot has been used here. The evidence for the reliability of the outcomes of scree plots for dimensionality is variable at best and therefore will be used here as a scoping exercise to obtain a general overview of the scale dimensionality before undertaking a more statistical analysis approach.

To assess the dimensionality of the data, a scree plot of the eigenvalues was graphed (Figure 6.0). The scree plot indicates that there is one dominant primary dimension as indicated by the steep drop between the first and second eigenvalues, a drop of approximately 23.6 points. However, there is another smaller drop between the second and third eigenvalues and again between the third and fourth eigenvalues before the plot levels off. This could be interpreted as either one or three factors/dimensions. Overall there appears to be one dominant primary dimension and two much smaller dimensions.

6.2.2 Analysis of the Residuals of PCA

As in Chapter 3 (section 3.6), the residuals of PCA of the Rasch model were analysed for the purposes of understanding the dimensionality of the scale. As before, the first step was to complete a preliminary analysis to study the person fit statistics to the Rasch model. The complete dataset for the 66 item pool set was analysed using RUMM2030 software. Where person fit statistics fell outside of reasonable limits (+/- 3.5) they were removed. The unidimensionality of the components was tested using the method as described by Smith (2002). The same method outlined in Chapter 3 (section 3.6) was adopted here to assess the unidimensionality of the overall crisis measure. Following the analysis approach outlined by Smith (2002), two subsets of items were compared using an independent t-test. If the measure is unidimensional then no more than 5% of the comparisons should be significant (Porta et al, 2011; Smith, 2002). Where <5% of comparisons are shown to be significant, this can be assigned to chance. When >5% comparisons are shown to be significant, the confidence intervals are wrapped around the statistic to account for the measurement error. When this occurs, unidimensionality is obtained where the binomial statistic is <0.05 and a binomial statistic of >0.05 indicates a bi-factor or multidimensional model.

Although possible unidimensionality was indicated in the analysis of the scree plot, the t-test following the PCA of the residuals indicated that >5% of the t-tests were significant, which is greater than that expected by chance. To check this result the confidence intervals were wrapped around the outcomes using the binomial test, which also indicated that the outcome did not fall within the confidence intervals and therefore the crisis measure is not unidimensional, indicating a possible bi-factor model or multidimensional model (binomial: no. < 5% = 73, N=254, Lower 95% CI-Proportion = 0.261 = >0.05).

Although the data suggests the presence of an overall general dominant factor, the statistical analysis did not show unidimensionality. Therefore it would not be appropriate to simply total the individual item scores to obtain an overall representation of the individual's need in terms of treatment from acute mental health services. A different approach to obtaining this overall impression of the individual's presentation will need to be developed (Chapter 7).

6.2.3 Hypotheses on the Structural Model of the Crisis Measure

The structural model of the crisis measure is not clear from the outcomes of these basic analyses. What is clear from the scree plot of the eigenvalues is that there is one particularly dominant primary dimension above a number of lesser dimensions. This may indicate that there is a dominant crisis dimension, e.g. Crisis Recovery Indicators, that influences the outcomes on the other lesser dimensions. For example, a high level of *cause for concern* on the Crisis Recovery Indicators may lead to increased levels of concern on the lesser subscales. Alternatively, it may be that there is a mediating dimension which acts as a filter through which the remaining dimensions are seen, for example, subscale 4: Mediating Factors focuses on the protective factors that support an individual to remain in the community and may well influence the *cause for concern* indicated on the other subscales such as Adaptive Decision Making and Risk of Harm to Self. A third hypothesis would be that the results may indicate a bi-factor model whereby the items provide information to a number of lesser dimensions as well as the dominant primary dimension. It does not appear that this is a purely multidimensional measure due to the demonstration of a particularly dominant primary dimension and therefore the hypothesis of a multidimensional model can be ruled out for future confirmatory statistical analyses through Structural Equation Modelling techniques.

6.3 Identification of Item Characteristics

The results of the scree plot indicated that there is one dominant primary dimension, but the outcome of the residuals of PCA analysis demonstrated that the item pool did not meet the criteria for unidimensionality with >5% of t-tests shown to be significant. Currently, there are very limited resources for the analysis of bi-factor graded response/rating scale models and those that are starting to become available are still being discussed in the literature and tested. The indication that there is one dominant primary dimension supported the decision to apply the data to the Rasch model for exploratory purposes. The Rasch model has been shown to be robust in estimating person and item parameters despite the violation of underlying assumptions such as the assumption of unidimensionality. This has been demonstrated through a number of simulated data research projects that specifically aimed to test the robustness of the Rasch model and other Item Response Theory models (Forsyth, Saisangjan & Gilmer, 1981; Harrison, 1986; Slinde & Linn, 1979b; Yang, 2007) and specifically that the violation of unidimensionality does not have an impact on item difficulty estimates (Yang, 2007). In a particularly useful model comparison Ip (2010) reported the results of a theoretical investigation into the empirical differences between multidimensional item response models (with a particular interest in models with one dominant primary dimension) and unidimensional models. The research concluded that the differences between these models were empirically indistinguishable with the outcomes representing the key dominant dimension, effectively describing the construct of interest. This understanding of measurement is much more in tune with the real world context where responses to items are generally influenced and/or determined by a number of factors.

The purpose of applying the dataset from this research to the Rasch model was to identify the items most representative of the key features of the item pool overall whilst acknowledging that there may be limitations in terms of interpretation due to the dataset not meeting the unidimensionality criteria.

The full 66 item dataset was applied to the Rasch model to identify the items that are most characteristic of the crisis assessment. Items that were shown to have the closest fit to the Rasch model may indicate key assessment items for crisis and assist the future development and refinement of the measure. In addition, the location of items on the Rasch logit scale was analysed to identify items that were less likely to receive a score and therefore may indicate higher levels of concern when scored, possibly providing key risk indicators for assessment.

6.3.1 Identifying Items Most Representative of the Crisis Measure – Item Fit Analysis

The item fit residual is the difference between the observed and Rasch expected item score (Bond & Fox, 2007; DeMars, 2010). The closer the item observed outcome is to the Rasch model expected outcome the closer to 0 the fit residual statistic will be. In reality, item observed scores rarely mirror exactly the Rasch model predicted outcomes. Items that have fit residuals close to the 0 point may represent the key features of crisis assessment and would be considered the assessment items that provide the most relevant information to the decisions made based on this model. Table 6.0 and Figure 6.1 outline the ten items with the smallest fit residual statistics, which identifies them as the most representative of the crisis measure's item pool. The items that comprise *ability to take control* and *intent to commit suicide* were shown to fit the Rasch model closest.

Table 6.0 – Fit Residuals for the Crisis Measure

Subscale	Item	Fit Residual
Recovery Indicators (1)	Ability to take control	-0.006
Risk of harm to self (3)	Intent to commit suicide	0.021
Adaptive Decision Making (2)	Response to hallucinations	0.065
Risk of Harm to Others (6)	Harm to others	0.070
Mediating Factors (4)	Support networks	0.086
Adaptive Decision Making (2)	Insight	-0.087
Basic Needs (8)	General wellbeing	0.111
Risk of Harm to Others (6)	Risk of neglect of others	0.135
Recovery Indicators (1)	Ability to relax	0.143
Mediating Factors (4)	Social Circumstances	-0.147

Table 6.0: indicating the fit residuals of the 10 items that fit the Rasch model closest for the crisis measures' 66 item pool.

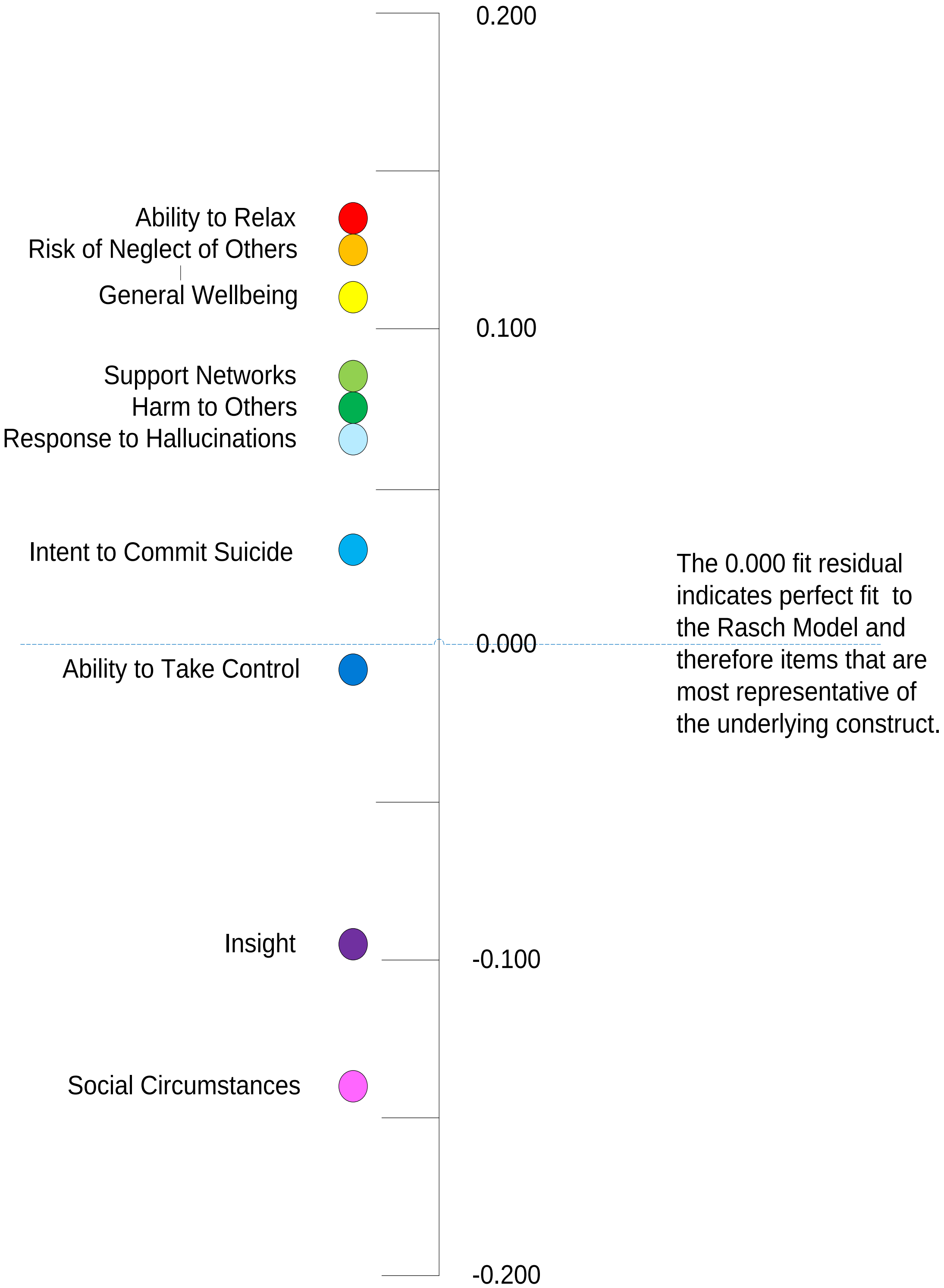


Figure 6.1 Plot outlining the 10 items that have the smallest fit residuals when their observed outcome was compared against the Rasch model' expected outcome. Items falling closest to the 0 point in the centre of the logit metric are most representative of the Rasch expected model.

The items shown to be the most representative of the underlying construct covered by the 66 item pool have been taken from all but 2 of the subscales. Subscale 5 for Daily Structure and Subscale 7 for Feelings & Affect are not represented by these 10 items. The 5 closest fitting items to the Rasch model will be discussed further here.

Interestingly, *ability to take control* is shown to be the most representative item for this item pool and therefore most representative of the core underlying construct for the measure. This item has strong links to coping theory, which is one of the underpinning theories of crisis (Chapter 1, section 1.0.4.4). The item *ability to take control* was conceptualised in the definition list for the crisis measure as a person's ability to manage situations effectively and to steer situations towards a desired outcome. The guidance for scoring this item emphasises an appraisal of a person's ability to effectively and helpfully manage and cope. When compared to the definition for crisis by Roberts (2002) as outlined in Chapter 1 (section, 1.0.2.3) and the working crisis definition of this research (Chapter 1, section 1.0.5) it can be seen that where an individual is presenting as a cause for concern on the item *ability to take control,* their usual coping mechanisms will have failed and they will perceive themselves as being unable to face stressors. This also fits in well with more recent research that has linked poor coping and social problem solving with suicidality (Pollock & Williams, 2001; 2004) and good coping and problem solving has been related to resilience (e.g. Grover et al., 2009; Priester & Clum, 1993b; Yang & Clum, 1994).

The following 3 items (*intent to commit suicide*, *response to hallucinations/delusions*, *risk of harm to self*) relate to the subscales 'risk of harm to self', adaptive decision making and risk of harm to others. Linking this back to crisis theory and its underpinnings these items appear to link with principles related to coping and self-management specifically relating to thoughts and cognitions and the potential behavioural consequences. The Yerkes-Dodson arousal curve (1908) outlines the relationship between arousal and performance (Chapter 1, section 1.0.2.3), suggesting that when arousal is significantly increased, the ability to cope and to be resourceful is hindered and as a consequence, performance is hindered. In the context of Decision Theory (Baron 2000) and the evidence indicating that judgement and decision making are significantly influenced by increased emotion and arousal, it is clear how increased arousal leading to and resulting in crisis would have an impact on performance in terms of decision making and resulting behaviours. The term performance or resulting behaviours described here relates directly to the items regarding an individual's *response to thoughts/hallucinations/delusions*, risk of harm to self in terms of their *intent to commit suicide* and *risk of harm to others*. Where routine coping fails, individuals experiencing crisis will often look for more extreme answers to their predicament. Under already pressured circumstances, these answers for their crisis experience are not always the most helpful possible resolution choices. Research has suggested that there are a number of cognitive biases in a range of processes such as memory and rumination which have been associated with increased likeliness of suicidality (Morrison & O'Connor, 2008; O'Connor & Noyce, 2008) and relates to the three items identified, which all have aspects of risk of harm to self and therefore risk of suicide. An individual's ability to manage and cope with their difficulties will be significantly impaired throughout the duration of crisis. Ultimately, the safety of the patient and of others in the community is of primary concern when treating an individual in their home environment where professional support cannot be continuous. Therefore, where safety and risk is of paramount importance, it is not surprising to see that *risk of suicide*, *response to hallucinations/delusions* and *risk of harm to others* demonstrate some of the closest fits to the Rasch model for crisis assessment and have been shown to be some of the most representative items of the underlying dominant construct for this measure.

Interestingly, *support networks* are indicated as the 5th most representative item for this measure. Hobbs (1984) indicated the importance of protective factors and vulnerability factors in the development of a crisis state (Chapter 1 Figure, 1.1), which is one of the underpinning theories in the crisis literature. This research takes this concept further, clearly indicating the importance of protective factors (mediating factors), specifically the patient's networks of support for managing crisis and working towards a helpful crisis resolution in the community. *Support networks* were described in the definition list for this measure as the relationships with other individuals and groups that support a person to function both physically and emotionally in the community. This outcome, suggesting that support networks are particularly important for successful home treatment, reflects a recent study that identified a person's perception of social support as the major predictor of lower levels of suicidal ideation independent of depression and hopelessness (Chioqueta & Stiles, 2007). Therefore, where support networks are indicated to be a particular cause for concern, more support from either the community or inpatient services would be required.

6.3.2 Item-Person Map

Figure 6.2 shows the RUMM2030 Item-Person map on a horizontal scale. The items are mapped out by their rating-scale categories. For example item 1 has 3 categories on its rating scale and each category point is mapped onto the item-person threshold map. This shows that overall the item difficulty (N=330, mean location is fixed at 0, SD = 1.114) is higher than the overall level of *cause for concern* presented by the sample (N=330, mean location = -2.512, SD = 1.377). When this is considered in the context of the population the crisis measure was piloted with, it is not surprising to find that the item difficulty has been shown to be higher, in terms of the location along the logit scale, compared to the general level of *cause for concern* posed by the sample. For the purposes of measuring crisis, the patterns in the layout/fall of the item and person difficulty levels are both expected and preferred. It is expected because, as outlined in Chapter 3 (section 3.3) the crisis sample represented on the crisis continuum from 'no crisis' through to 'inpatient admission' whereas the item pool was developed to represent only the 'crisis present' part of the crisis spectrum. Therefore, there will be a number of patients toward the less acute end of the crisis spectrum, not considered to be in a state of crisis and therefore not taken on by the CRHT. This was also indicated by the descriptive statistics outlined for the individual subscales (Chapter 5, section 5.4) where the data was skewed to the zero point as shown by the Shapiro-Wilkes outcomes.

Figure 6.2- Item-Person Threshold Map for the CRAFT

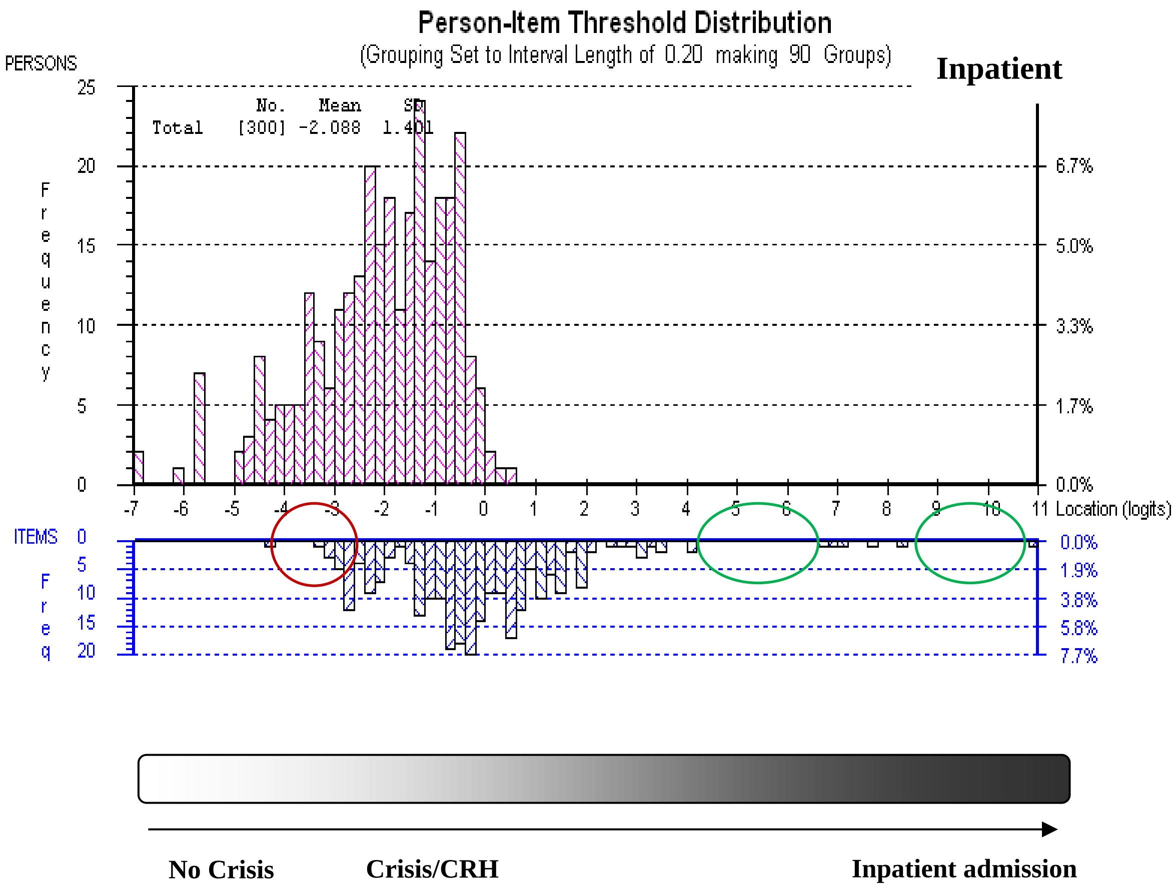


Figure 6.2: Item-Person Threshold Map (item thresholds/categories are represented in blue and persons represented in red) outlining the spread of the items against the sample distribution. Below the Item-Person map is the hypothesised relationship between the Item-Person spread and a crisis spectrum.

The aim of scale development is for the measure to have the capacity to represent the crisis construct at each level of the crisis spectrum. For the purposes of CRHT assessment it is particularly important that the measure can differentiate between those who present with crisis and those who do not. As the measure was piloted to represent all crisis levels from no crisis through to inpatient admission, the data collection has been used for this purpose. However, the items were developed from interviews and focus groups that specifically concentrated on indicators of the crisis state and therefore it would be expected that items would receive scores indicating *cause for concern* once the crisis state was shown to be present. Based on this understanding, it is hypothesised that the item difficulty will start at a higher level on the Rasch location metric than the person levels due to the items starting at the point where crisis is indicated (circled on Figure 6.2 in red), whereas the crisis sample will represent both the non-crisis and crisis populations on the Rasch location spectrum.

It was shown that there are two breaks in the spread of the items towards the higher end of the location spectrum (circled in green in Figure 6.2 above). There is one item, *speech*, that is shown to be standing alone at >+11 logit area on the location metric, which infers that this is a particularly unlikely item category to receive a score for, i.e. to have significant cause for concern is one of the probable response categories. On closer inspection of the Item-Person Category location map (Appendix 13) this item category was shown to have a rare occurrence and its isolation on the map suggests that for an individual to show this level of difficulties with their speech and/or communication of need would present as a significant concern and a risk indicator. Considering the context of the CRHT providing treatment in the community, it is logical to expect an individual's ability to effectively communicate and to have good speech would be vital for home treatment. An individual's risks will increase where that person is unable to communicate their needs effectively as discussed more fully in the previous chapter (section 5.7).

After this item category (Figure 6.3) there is approximately a 6 logit gap before a clustering of 13 item categories around the 5 logit to 8 logit locations on the metric (Appendix 13). The next highest item categories on the item rating scales were *appropriateness of mood*, *domestic violence*, *risk of neglect of others* and *poverty of thought.* The remainder of the item category locations fell between the 4 logit and -4.5 logit locations. The item *ability to manage symptoms* was shown to be the most likely item to receive a score for on the first category of the *cause for concern* rating scale. Again, this links in with the working definition presented by this research (Chapter 1, section 1.0.5), focusing on a person's ability to manage and cope with their difficulties in order to function in an adaptive manner in the community and to resolve their crisis in a helpful way. Ability to manage symptoms also links in with the idea that crisis significantly impacts a person's ability to make adaptive decisions and to be resourceful, ultimately impacting on the person's ability to make decisions, often shown through their behaviour and speech.

The data for analysis was obtained for a community crisis population and it is seen that the majority of people and items group together centrally and tail off towards either side approximately following a normal distribution. It could be deduced that the bulk of the central tendency represents those falling within the community crisis population with those falling towards the tails representing the extremes of the sample population i.e. those not in crisis towards the left hand side of the scale and those with particularly acute presentations and possibly requiring admissions towards the right hand side of the scale. The hypothesised relationship between the Item-Person map and the crisis spectrum is indicated below the item-person map in Figure 6.3.

6.3.4 Item Risk Indicators

Items that are shown to be less frequently scored through Rasch analysis are estimated by anchoring either the sample ability or the item difficulty, i.e. at a mean of 0 with a standard deviation of 1. The item difficulty is the level of *cause for concern* required for approximately 50% of individuals to be rated on that item. Therefore, if an item has a difficulty estimate of 0.2 then individuals with a *cause for concern* level at 0.2 would be expected to rate on this item 50% of the time. Figure 6.3 below is the Item-Person location map on a vertical scale. This map indicates the individual items (items as a whole without differentiating between the rating scale categories) and their locations along the logit scale. The items toward the top of the scale (positive logit scale numbers) indicate items that are the least likely to receive a rating for as a whole. The items toward the bottom of the scale (toward the minus logit numbers) indicate the items that are most likely to receive rating. Similarly, individuals shown to be toward the top of the metric are those whose *cause for concern* is greater and therefore they are more likely to be experiencing crisis. Examination of the item location tables on RUMM2030 supports accurate identification of item and person locations. The 10 items least likely to receive ratings are outlined in Table 6.1 below.

Figure 6.3 Person-Item Map Outlining Item Difficulty Locations

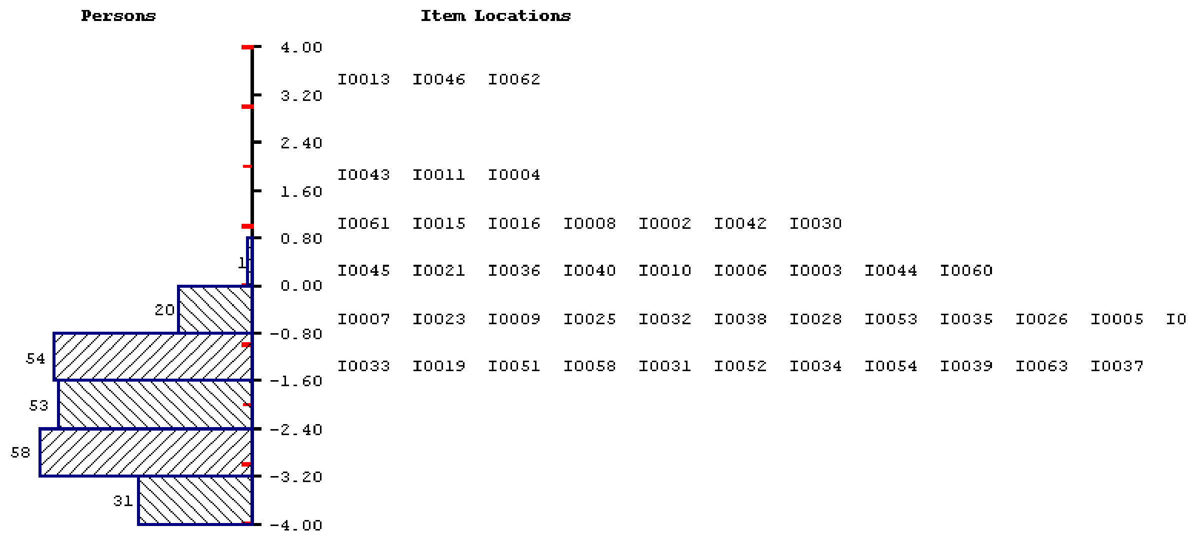


Figure 6.3 Example of Item-Person location map indicating the spread of item to person difficulty estimates. Items (on the right hand side of the metric) toward the top of the metric are indicated as the most difficult. Not all items are displayed as they run off to the right of the Figure.

Key: Item key – items are numbered according to original 143 item pool for RUMM2030 analysis. The table below indicates the item labels for the top three rows of items:

Item Number in Figure 6.3	Item Label	Item number in final 66 item crisis measure
10062	Appropriateness of mood	66
10046	Domestic violence	57
10013	Speech	29
10004	Poverty of thought	20
10011	Overall acceptance of support	27
10043	Risk of neglect of others	54
10030	Ability to take control	16
10042	Violence and aggression	53
10002	Stream of thought	18
10061	Appetite	65
10015	Overall thought content and clarity	1
10016	Concentration and attention	2
10008	Confusion	24

Table 6.1 – Items Locations for the Items Least Likely to be Rated

Subscale	Item	Item location
Basic Needs (8)	Appropriateness of mood	3.639
Risk of Harm to Others (6)	Domestic violence	3.517
Adaptive Decision Making (2)	Speech	3.295
Adaptive Decision Making (2)	Poverty of thought	2.137
Adaptive Decision Making (2)	Overall acceptance of support	2.025
Risk of Harm to Others (6)	Risk of neglect to others	1.810
Recovery Indicators (1)	Ability to take control	1.279
Harm to Others (6)	Violence and aggression	1.256
Adaptive Decision Making (2)	Stream of thought	1.223
Adaptive Decision Making (2)	Capacity to consent	1.165

Table 6.1 Table indicating the logit locations of the 10 items that are shown to be most difficult to rate based on the Rasch model for the 66 item pool.

The items outlined to be the least likely to cause concern are found in subscales 1- Recovery Indicators, 2 – Adaptive Decision Making , 6 – Risk of Harm to Others and 8 – Basic Needs. The majority of the items appear to centre around thought processing and harm, which may suggest that these are key risk indicators for crisis assessment. The top five items suggested through Rasch analysis to be helpful risk indicators will be described in more detail below.

Although Rasch analysis can draw attention toward those items that are shown to be particularly difficult to receive ratings, it cannot indicate the reasons for why this has occurred. It is therefore up to the researcher to make these inferences or hypotheses. *Domestic violence* is shown to be the second most difficult item to receive a rating for *cause for concern*. It could be considered that this is because domestic violence only occurs in the most serious incidences of crisis and therefore when rated indicates someone who is particularly at risk and requires more structured support. However, it may simply reflect the general levels of incidence observed for domestic violence in the population/community as a whole rather than as a particular risk indicator for the crisis population.

The difficulty for this research is that the base rates of occurrence for the crisis population are not known and therefore there is no appropriate context in which to judge the difficulty of items. In addition, items that are shown to be difficult to rate as a *cause for concern* could be specific to a particular condition or presentation. For example, scoring for *cause for concern* on the item 'speech' may be specific to some particular types of presentation, e.g. acute psychotic episodes. Alternatively, items that are shown to be less likely to cause concern may act as particular indicators of risk, and therefore obtaining a score on these items, even a low one, may indicate particular *cause for concern*.

It is important to interpret the information provided from the difficulty estimates flexibly and with caution. Although items that are indicated as particularly 'difficult' may act as red flags for assessment, it should be appreciated that some of these items may simply be difficult due to the frequency of the under-reporting for example, rather than them being particular indicators of risk.

Appropriateness of mood was shown to be the item least likely to be rated as causing concern through Rasch analysis. This item is supported clinically through previous identification of this item as a key indicator of mental state as demonstrated through its use in both The Psychiatric Interview and Mental State Examination, which are both used worldwide (Andrews & Ovsiew, 2009).

Domestic Violence and *speech* were both identified as risk indicators for their subscales and therefore won't be discussed further here as they have already been discussed in relation to crisis assessment previously (Chapter 5, section 5.8).

Poverty of Thought describes a significant breakdown or block in an individual's thought processing which is often observed as an inability to answer questions, stopping mid-sentence or moving across topics within a conversation that does not flow or follow. This item is used as part of diagnostic assessment for Schizophrenia or psychotic experiences and is also included in The Mental State Examination (Akiskal, 2008). As well as providing a useful symptom criterion for assessment on a practical level this also relates to some of the considerations made for concerns related to the item *Speech.* Communication is particularly important for home treatment and particularly for individuals who are experiencing mental distress at levels as acute as those associated with mental health crisis. Therefore, difficulties in communication associated with *poverty of thought* would be particularly problematic for treating a patient in their home environment and would present a particular risk indicator for crisis home treatment.

Overall Acceptance of Support reflects an individual's willingness to engage and accept treatment from the CRHT but also encapsulates the ability of the patient to engage in support provided from their closer support networks through family, friends and the community. The importance of engagement and treatment adherence has been highlighted by a number of authors in the field of crisis (e.g. Tacchi & Scott, 2008), acknowledging the collaborative approach essential to community treatment approaches.

6.4 Appraising the Usefulness of Whole Scale Cut-offs

The evidence presented in this Chapter of the research suggests that the crisis measure does not meet the criteria for statistical unidimensionality although there is evidence of one primary dominant dimension from the scree plot (Figure 6.0). This presents a challenge to the decision to sum the item scores to provide a total. The Rasch literature suggests that unidimensionality should be assessed through a number of approaches both statistically and using methods such as scree plots. As expected the measure did not meet the criteria for statistical unidimensionality which was suggested by the identification of 8 subscales through the PCA. However, there is a strong indication that there is a primary dominant concept that the items contribute information to and ties these subscales together. From a theoretical perspective this would also be expected as the items are all being scored within a very specific context of crisis assessment and therefore the way the items are scored will be contributing information to an overarching understanding of the context. The specific point of interest here is to try and understand what this specific context really is. Is it as simple as crisis assessment or is it some other overarching concept such as '*community treatment*' or '*ability to effectively work* *with CRHT'* for example. This indicates that summing the item scores to provide an overall total score for this measure would not be a legitimate representation of a level on a core underlying construct due to evidence that the items may be representing more than one key dimension. However, this does not mean that a whole scale cut-off cannot be provided for this measure using other approaches. These will be explored further in Chapter 7.

6.5 Summary and Conclusions

The evidence provided from this analysis for the crisis measure indicates that it is most likely to meet the criteria for a bi-factor model. What this suggests is that the information provided by the items on this scale informs an overall picture of a crisis construct as well as more specific information about focused areas that contribute to the overall understanding of the crisis presentation. In terms of utility, the bi-factor model is particularly helpful for providing useful information to CRHT mental health professionals both in terms of providing an overall crisis picture but also for informing treatment plans by identifying particular areas of strength and weakness.

The indication of a bi-factor model would suggest that there is a core underpinning dimension that brings harmony to the overall scale, giving the subscales a common ground on which to function. The challenge here is to understand what information the underlying, underpinning dimension offers to the assessor. Is it an overview of the construct of acute mental health crisis or is it some other construct integral to the assessment of individuals who are experiencing crisis? This hypothesis could be further explored using techniques embedded in Structural Equation Modelling in future research. At this point however, it is helpful to draw together the evidence to guide toward the most appropriate label for the construct represented by this scale. For example, the evidence obtained in the interview and focus group stage of this research (Chapter 2) suggested that it was not the symptom level that was important for crisis assessment but the level of *cause for concern* in terms of that person's ability to manage and cope in a healthy adaptive way, working towards a positive resolution of their crisis.

This information was supported by an appreciation of both the risk and protective factors relating to each item which helped to determine how much of a cause for concern that person was. In terms of the evidence provided from this analysis, it was shown that the items most representative of the measure's underpinning construct tap into concern for an individual's ability to cope and manage at home based on their own personal strengths, the support of others, and their risk of harm to self and others.

The evidence appears to suggest a theme relating to an individual's ability to cope in a healthy way, with appropriate support structures, in order to manage in their own environment and to function in an adaptive manner appropriate for the community setting. Where the cause for concern is raised it indicates some impairment of that person's ability to manage in their own environment, to function in the community setting in an adaptive way and to stay safe. Ultimately, this is represented by a decrease in protective factors and an increase in risk factors in the 8 key areas identified through this research. *Adaptive Community Functioning* is the phrase chosen to describe this underpinning core theme or concept described for this measure. This term brings together the key items that represent the Rasch model closest and is suggested to best describe the core concept held by this model. Where an individual is shown to be able to function in their community in an adaptive manner, using appropriate coping strategies, support networks and being able to keep themselves and others safe, the requirement for CRHT support is minimal. Where an individual is not shown to be functioning in the community in an adaptive manner but to be struggling to cope and manage, isolated from their community and ultimately placing that person or others in danger of harm indicates the necessity for support from the CRHT if not inpatient services is clear. Based on this understanding of the underlying core concept to this crisis measure, the measure was renamed as the Crisis Risk and Adaptive Functioning Tool (CRAFT). The 66 item CRAFT is outlined in Appendix 18.

Having gained more clarity regarding the underpinning concept for these items it is possible to appreciate more fully the potential utility of providing total score outcomes for this measure. An overall outcome score may be able to provide useful information about the overall ability of an individual to function in an adaptive manner in the community. Although unidimensionality has not been demonstrated statistically, if a scoring model can be developed to meet acceptable levels of sensitivity and specificity it would be able to offer valuable guidance to mental health professionals. The next step in this research is to look at possible scoring models and their ability to provide meaningful information and outcomes.

Chapter 7

Crisis Risk and Adaptive Functioning Tool

Treatment Indicators of Crisis

7.1 Introduction

The eight key assessment areas for CRHT teams have been outlined (Chapter 3) and the key indicators for each subscale identified through Rasch analysis (Chapter 6). Cut-offs for each subscale were developed using percentile ranges of the interval level Rasch logit scale (a scale with equal distances between categories as defined in Chapter 1, section 1.3.1.2) to represent very low, low, moderate, and high levels of *cause for concern*. The next step is to identify whether the summed total for the overall Crisis Risk and Adaptive Functioning Tool (CRAFT, Appendix 18) measure can provide a meaningful outcome pattern that relates to treatment decisions made in clinical practice in order to develop helpful and informative cut offs.

It is the subtle interaction of a person's areas of strength and weakness across the elements of the 8 subscales and how they weave and link together to form the overall picture of crisis which is critical for an overall understanding of an individual's experience. Understanding this complete picture helps the CRHT to assess where the patient is on the treatment spectrum ranging from not requiring CRHT treatment through to requiring home treatment supported by the CRHT or care delivered through inpatient services.

It was suggested in Chapter 6 that the underpinning dominant dimension of the item pool for the Crisis Risks and Adaptive Functioning Tool (CRAFT) is *Adaptive Community Functioning*. It is therefore expected that the overall picture created from combining the information from the subscales should provide an image of a person's ability to function adaptively in their environment in order to function in the community. This information is useful for CRHTs who ultimately act as gatekeepers to inpatient care (National Service Framework, 1999), making it essential that they are able to accurately understand a person's ability to function safely in the community. The CRHT team need to take into consideration all aspects of the patient's presentation in order to inform appropriate treatment decisions. Where an individual demonstrates an inability to adaptively function in the community, this will provide helpful information to support the decision to provide treatment through inpatient services rather than in the community.

The next step for this research was to identify if there is a pattern to the summed CRAFT total score that will reliably indicate whether or not an individual requires support from the CRHT for managing crisis and ultimately to indicate the level of treatment required, related to the 5 treatment levels currently used as part of standard practice indicating discharge/low (green), moderate (amber), high (red) and admission.

Comparing the pattern of the overall total outcome scores to the actual outcomes indicated by the team at the time of assessment will help to clarify whether there is a relationship between the CRAFT total score and the treatment decision made at the time of assessment. This supports the identification of clinically credible and meaningful cut-offs to support CRHT treatment decisions which can then be tested using analysis of the cut-off's sensitivity and specificity. There is no current gold standard for crisis measures in practice and therefore this measure will attempt to mirror what is indicated in the field (criterion validity) for this first stage of the measure's development.

The subscales of this measure have been developed to meet the criteria for unidimensionality (Chapter 3, section 3.6). It can therefore be assumed that the summing of the item scores for the subscales is measuring one underlying construct providing a meaningful indicator of that construct level (Pallant & Tennant 2007; Thurstone, 1931). In order to justify the addition of the subscale scores to provide a total score for the whole measure, the overall scale was also assessed for unidimensionality as well. Although the scree plot indicated that there was one dominant primary dimension this was not supported by the statistical analysis. The evidence provided from these analyses (Chapter 6) may indicate a bi-factor model that taps into both one dominant underlying construct in addition to contributing information to a number of smaller dimensions (Reise, Morizot & Hays, 2007).

Although it may not be legitimate to simply total the item scores based on the outcomes of the statistical assessment of unidimensionality, there may be an alternative approach to combining the information provided by the subscales in a valid manner. It has been recognised throughout this research that measurement development is a delicate balance of theoretical, statistical and practical considerations. Where the outcome score of the global overall CRAFT measurement tool can be shown to be meaningful, providing acceptable levels of sensitivity and specificity, the clinical and practical advantages of this may far outweigh the theoretical drawbacks. This would legitimise this step in the research.

The aim of this section is to explore a valid and reliable method for providing an overall impression of the crisis presentation to guide, support and improve CRHT treatment decisions:

1) Explore an alternative valid method for totalling the CRAFT subscale scores
2) Develop scale total scoring models for providing potential whole scale cut offs
3) Analyse the scoring models developed for sensitivity and specificity
4) Choose the scoring model most appropriate for supporting CRHT assessment.

7.2 The Crisis Measure Total Score

In order to provide an overall scale outcome for this measure, data from each of the subscales is combined to provide an overall impression of the CRAFT outcomes. However, it is not possible to simply add the subtotals together due to the statistical evidence suggesting that the CRAFT is not unidimensional and, on a more logical basis, the subscale totals are all on individual scales that differ in length from 13 to 61. For example, subscale 2 has a total of 61 and Subscale 7 has a total out of 13 and therefore issues around ‘weighting’ need to be taken into consideration and accounted for in developing a whole scale scoring model.

The use of percentiles to divide the subscale totals into levels of very low, low, moderate, and high allows the subscales to be compared on an equal platform out of a possible total of 4. The percentiles were calculated from the interval level logit scale and so may be compared in terms of looking at where the areas of strength and weakness lie.

To obtain a total score, the percentiles were given a rank from 0 to 4 (0=0, 1=0-20th percentile, 2 = 21-40th percentile, 3 = 41-60th percentile, 4 = 61st+ percentile) and a scoring table developed to support CRHT practitioners to transform the raw score into their equivalent rank (Appendix 10B). The ranks were added together to give a total score out of 32. The decision to transform the percentiles into ranks rather than simply retaining the percentile was purely for practical reasons to support the mental health worker to complete the scale by hand when necessary. The mental health worker would not be subjected to the transformation calculations that occur between the raw score and the final rank as these have already been calculated (Figure 7.0, step 4) and are captured in the scoring tables completed through this research (Appendix 10B). Where computer software is developed, the raw score will be automatically transformed into an outcome indicator. Figure 7.0 outlines the steps taken to obtain the final subscale percentile rank outcome which will then be totalled for the overall CRAFT total. The total scale score may then be used to indicate the level of treatment required where an appropriate scoring model with adequate sensitivity and specificity is developed.

Figure 7.0- CRAFT Score Transformations

Step 1: The assessor indicates the raw score on the item scale. The score here is underlined and highlighted.

<table>
<tr><td>Item 1</td><td colspan="4">Cause for Concern</td><td colspan="3">Neutral</td><td colspan="4">Not Cause for Concern</td></tr>
<tr><td>Raw Score</td><td>5</td><td>4</td><td><u>3</u></td><td>2</td><td>1</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
</table>

Step 2:

The scoring template overlay is used to access the transformed item scale view (Appendix 15). The new position of the score indicated in step 1 with an underline and the box highlighted. This indicates that the rating indicated is equivalent to a score of 3.

<table>
<tr><td>Item 1</td><td colspan="4">Cause for Concern</td><td colspan="3">Neutral</td><td colspan="4">Not Cause for Concern</td></tr>
<tr><td>Raw Score</td><td>5</td><td>4</td><td><u>3</u></td><td>2</td><td>1</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>Template Score</td><td colspan="2">3</td><td colspan="2">2</td><td colspan="3">1</td><td colspan="4">0</td></tr>
</table>

Step 3:

This item is part of a subscale of 4 items and therefore the 4 scores are added together from the transformed item scale template. This gives a total of 7 for this subscale.

<table>
<tr><td></td><td colspan="4">Cause for Concern</td><td colspan="3">Neutral</td><td colspan="4">Not Cause for Concern</td></tr>
<tr><td>Item no.</td><td>5</td><td>4</td><td>3</td><td>2</td><td>1</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>1</td><td colspan="2">3</td><td colspan="2"><u>2</u></td><td colspan="3">1</td><td colspan="4">0</td></tr>
<tr><td>2</td><td colspan="2">3</td><td colspan="3">2</td><td colspan="5"><u>1</u></td><td>0</td></tr>
<tr><td>3</td><td colspan="2"><u>3</u></td><td>2</td><td colspan="7">1</td><td>0</td></tr>
<tr><td>4</td><td colspan="2">4</td><td>3</td><td>2</td><td colspan="3"><u>1</u></td><td colspan="4">0</td></tr>
</table>

Step 4:

The example subscale is then transformed into the Rasch logit scale which represents an interval level scale. The subscale total of 7 is equivalent to a logit scale score of 0.222 in this example. This is then assigned to a percentile band which is ranked. This is not a linear relationship and therefore the steps in the process must be followed to obtain an accurate representation of rank level. The columns in grey indicate the transformation stages that are not visible to the assessor. As indicated by the arc arrow at the top of the table, the assessor could simply transform their score into a rank.

Subscale total score	Subscale score transformed into the logit interval level scale	Percentiles calculated based on the logit scale	**Percentile ranked 0-4**
0 1 2 3 4 5 6 **7** 8 9 10 11 12 13	-4.789 -3.715 -2.860 -2.180 -1.563 -0.969 -0.384 **0.222** 0.893 1.596 2.263 2.948 3.783 4.803	0 0 1 20th 2 40th 3 60th 4	0 1 2 **3** 4

Step 5:

The table below provides an easy method for transforming the total subscale score into the final percentile ranked score (complete set of tables is in Appendix 10B). The rank scores for each subscale are then totalled to give the overall CRAFT total score:

<table>
<tr><th>Raw</th><th>Percentile Rank</th></tr>
<tr><td>0</td><td>0</td></tr>
<tr><td>1</td><td rowspan="4">2</td></tr>
<tr><td>2</td></tr>
<tr><td>3</td></tr>
<tr><td>4</td></tr>
<tr><td>5</td><td rowspan="4">3</td></tr>
<tr><td>6</td></tr>
<tr><td>7</td></tr>
<tr><td>8</td></tr>
<tr><td>9</td><td rowspan="5">4</td></tr>
<tr><td>10</td></tr>
<tr><td>11</td></tr>
<tr><td>12</td></tr>
<tr><td>13</td></tr>
</table>

7.3 Data Preparation and Initial Analysis

Using SPSS the individual item scales were transformed into new variables with the Rasch refined item rating-scale scores. These item scores were totalled for each subscale and the subscale total transformed into the Rasch interval logit scale. The total logit score was then placed on the 4 point percentile scale and ranked. The subscale percentile ranked scores (on the 4 point scale) were totalled for the global overall measure score out of a possible 32. A histogram of the CRAFT total scores and basic descriptive information is outlined in Appendix 16. This shows the CRAFT total scores to fall approximating a normal distribution with approximately 84% of the total score scale being actively used with a range of 27 out of a possible 32 maximum total by this sample with an SD of 6.11 points. There were 71 missing values, which is accounted for due to the original pilot crisis measure containing 143 items and the frequency that items were either missed or not completed. This in itself probably suggests that the original measure was too long and support for the item reduction approach of PCA used in the first analyses.

Receiver Operator Characteristic (ROC) curves are graphs that map the sensitivity of a measure against the specificity of the measure (Altman and Bland, 1994). The sensitivity of this measure is the proportion of patients who require CRHT support who are correctly identified as requiring support by the measure. The specificity of this measure is the proportion of patients who do not require CRHT support and who are correctly identified as not needing CRHT support (Berwick, Cheek & Ball, 2004). A perfect test would have a sensitivity and specificity equal to 1. If a measure is unable to produce any meaningful cut-offs then it would be equally likely to produce a false positive (falsely indicating a need for treatment) as a true positive (accurately indicating a need for treatment). By mapping this information on a graph it is easy to see by eye how well a measure is able to differentiate between the treatment group (requiring CRHT support) and the normal group (do not require CRHT support). If a measure operates at a completely random level (i.e. no better than guessing) then the data will follow the pattern of a straight line from bottom left to top right of the graph (this is indicated on the ROC curve output in SPSS). Where the measure has perfect discrimination between the treatment group and normal population group there should be a curve going toward the top left side of the graph. In reality perfect discrimination between the groups rarely happens and overlap is often observed especially in measures designed for mental health. To statistically test the ability of the measure to discriminate between the two groups, the Area Under the Curve (AUC) is assessed. The ideal is for an AUC = 1 whereas a random guess would = 0.5. Therefore any AUC > 0.5 suggests that the measure is able to determine caseness better than random guessing (taking into account the standard error). Therefore the greater AUC, the greater the accuracy of the instrument.

7.3.1 Receiver Operator Characteristic Curve Analysis of the CRAFT Measure.

An initial ROC curve analysis was completed to assess if individuals requiring CRHT support were more likely to achieve higher scores on the crisis measure than individuals who did not require

CRHT support. Individuals who were indicated by the CRHT at the time of completion of the CRAFT to be an Amber, Red or Admission status were taken to be those requiring CRHT support. Individuals who were indicated by the CRHT at the time of completion of the CRAFT to be a Green or Discharge status and not assessed as being in a state of crisis were taken to be those not requiring CRHT support. This was assessed initially from studying the ROC graph (Figure 7.1) and from the area under the curve on the ROC curve (Table 7.0). The ROC curve was outlined by plotting the sensitivity of the crisis measure against 1-specificity of the measure and demonstrated a curve going toward the top left side of the graph, which suggests that the crisis measure is better at discriminating between the treatment and no-treatment group than chance (random guessing is indicated by the straight diagonal line). The AUC was assessed and it was shown that there is a greater probability of a person in the treatment group receiving a higher rating on the CRAFT than a person who is not in the treatment group (N=297, Area under the curve = 0.804, SE 0.027 lower bound 0.752 and upper bound 0.856). With the AUC indicated at 0.804 there is very good evidence to suggest that a scoring model can differentiate between individuals requiring CRHT intervention and those not requiring CRHT intervention and would be far better than random guessing. It is particularly encouraging to see such strong evidence at the early stages of measurement development for a mental health measure. The Hospital Anxiety and Depression Scale (HADS) has been shown to have an AUC of 0.81 for the depression subscale and 0.70 for the anxiety subscale (Bambauer, Locke, Aupont, Mullan & McLaughlin, 2005) when the outcomes of the HADS were compared to the Diagnostic and Statistical Manual of Mental Disorders (DSM IV) criteria. Hosmer and Lemeshow (2000) suggest areas under the ROC curve of 0.70 to 0.80 are 'acceptable', 0.80 to 0.90 'excellent' and 0.9 or above 'outstanding'. They point out ROC of 0.50 suggests no discrimination between the outcome groups as this corresponds to chance, e.g. simply tossing a coin to decide group membership. Therefore, the CRAFT measure falls in the ‘excellent’ category for the AUC.

Such a strong indication of the crisis measure’s ability to differentiate between the treatment and no-treatment crisis states suggest that it is a clinically valid option to develop a scoring model to differentiate between the different treatment levels.

Figure 7.1 Receiver Operator Characteristic Curve for the CRAFT

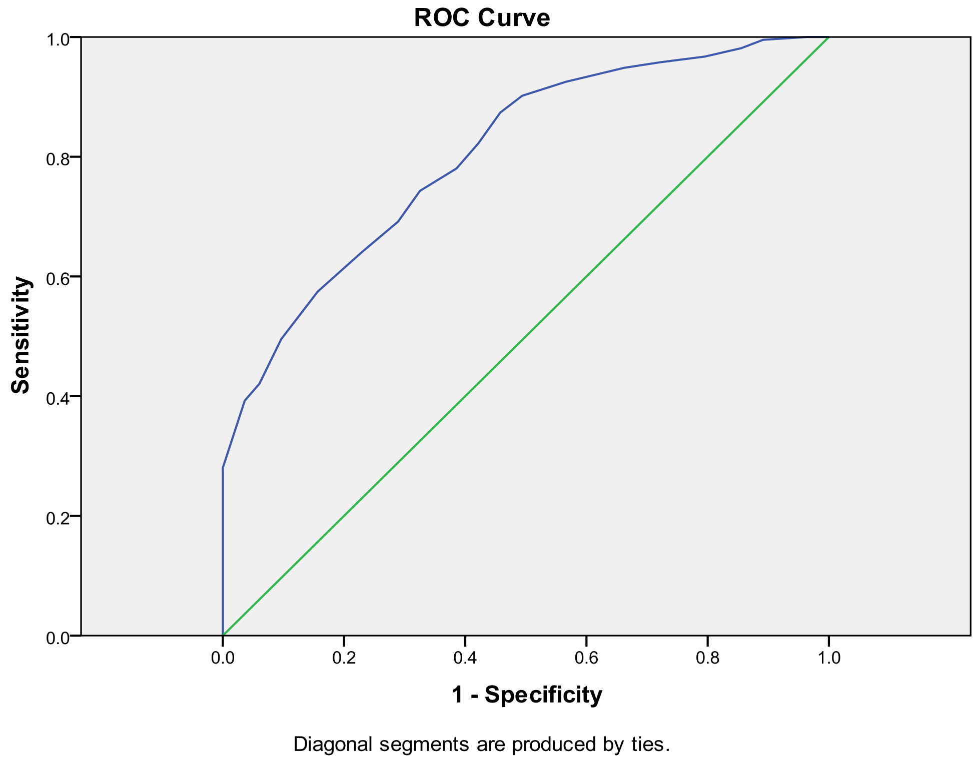


Figure 7.1: Receiver Operator Characteristic (ROC) curve for the CRAFT to indicate the ability of the measure to differentiate between individuals experiencing crisis and the non-crisis population. The ROC curve clearly shows an ability to differentiate between the crisis and non-crisis populations better than by chance. This is demonstrated by the ROC clearly curving up towards the left hand side of the plot. Curves that are closer to the central green line indicate an ability to discriminate that is no better than chance or less than chance if the curve goes down towards the bottom right hand side of the plot.

Table 7.0 CRAFT Area Under the Curve

Area	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.804	.027	.000	.752	.856

Table 7.0 Outcomes of the Receiver Operator Characteristic curve analysis indicating that a person experiencing crisis is much more likely (greater than chance) to receive a higher rating than on the crisis measure than a person not in crisis.

7.4 Scoring Model Development

The subscale percentile cut-offs identified in Chapter 5 were used to place all subscale outcomes on the 4 level ranking scale (representative of the percentile levels) indicated in Figure 7.0 (step 5). The crisis measure's overall scale score was calculated by adding together the subscales' ranked outcomes.

The first step in developing scoring models for the crisis measure was to study the spread of total scores obtained. From studying the total score spread it was identified that the highest total score achieved was 27 out of a possible 32 and therefore 15.6% of the total score scale was not actively used in this sample. As a result, the scoring models for this measure were developed based on the 27 point range total actively used by the sample. Based on this, three scoring models were developed (Figure 7.3).

The first model was developed as a control model by dividing the maximum obtained score of 27 into the 4 levels – these 4 levels have been labelled as very low, low, high and very high. These levels were then assumed to directly relate to the four levels of treatment outcome – discharge/Green, Amber, Red and Admission. There are a number of theoretical flaws due to the assumptions made regarding the scales' relationship with the treatment outcomes but it is functional for the purposes of a control model to act as a point of comparison for the other two models. The main theoretical difficulty with this model is the assumption that the 4 levels indicated by the division of the crisis measure's total score scale directly relate to 4 levels of treatment offered by the team. There is no statistical evidence to suggest this at this stage and therefore this scoring model is based purely on guessing and inference rather than on evidence.

The second model was developed using on information provided from Receiver Operator Characteristic (ROC) curves for each of the five assessment outcome levels of discharge, green, amber, red and admission (Appendix 17).

The most important aspect in developing this measure is the ability to identify individuals appropriate for CRHT support and therefore the focus for developing the second model using ROC curves was on the measure's sensitivity. The upper cut-off used for each level was a sensitivity of approximately 0.50 (50%) for the initial inspection (Table 7.1). This indicates the score at which 0.50 of that category will be correctly identified as meeting the requirements of that treatment group when an individual's score meets this cut-off. Therefore sensitivity increases as the cut-off score decreases. However, as sensitivity increases, the specificity of the measure often decreases and therefore a balance between the two has to be found.

Table 7.1: Receiver Operator Characteristic Curve Statistics

Treatment category	Receiver Operator Characteristic curve sensitivity statistic	Receiver Operator Characteristic curve Specificity Statistic	Receiver Operator Characteristic curve score	Total Score Cut off rounded up to nearest whole number
Admission	.491	.078	20.5	≤20
Red	.512	.333	16.5	≤16
Amber	.481	.614	13.5	≤13
Green	.481	.901	8.50	≤8
Discharge	.518	.896	7.5	≤7

Table 7.1 Receiver Operator Characteristic curve statistics for the five treatment categories used by the Bedford and Luton CRHTs. The sensitivity cut-off for approximately 0.5 (50%) was chosen as the primary cut-off for each treatment category and the associated score identified.

It was noted that the Discharge and Green categories demonstrated quite a lot of similarity and overlap in their ROC statistics (Appendix 17), as did the Amber and Red categories. When the Red category was divided into the lower (one visit per day) and upper (up to three visits per day or admission to the assessment unit) red categories, the upper red category was shown to have overlap with the admission category. On the basis of the above, it was decided to combine these categories together essentially developing 3 categories within the scoring model of:

1. No or low levels of CRHT intervention required (Discharge category + green category)
2. CRHT intervention required (Amber category + lower red category)
3. Intensive CRHT home treatment intervention or admission required (Upper red category + admission category).

The upper cut-offs for these new categories were identified from the ROC analysis and are outlined in column 2 of Table 7.2. It was decided to ensure overlap between the categories which would represent transition from one category to the next. It was felt that this would be more representative of what occurs in clinical practice. The upper cut offs for the discharge/green and amber/red categories were simply set at the lower cut-off of the category above, for example, the upper red/admission category lower cut-off starts at 21 and therefore the upper cut-off for the amber/red category was set at 21. The lower cut-off for the upper red/admission and amber/red categories was set by meeting the previous category’s higher cut-off for example, the amber/red category’s lower cut off was set at 8 which was shown to be the Green category’s higher cut-off on the ROC analysis. This resulted in a 4 point overlap between the categories, which reflects a more clinically relevant continuum model with areas of overlap representing movement between the treatment areas. The continuum model supports the concept that wellness and illness run along a spectrum rather than being distinct categories (Keyes, 2002; 2005; 2007) and reflects current thinking in relation to diagnoses. This concept for the scoring system and the idea of crisis functioning on a spectrum has been outlined in Figure 7.2 below.

Table 7.2 CRAFT scoring categories based on ROC analysis outcomes (Model 2)

Treatment category	Total Score Cut off indicated at approximately 0.5	Categories	Score ranges
Admission	≤20	3) Inpatient admission or intensive CRHT intervention.	≥17
Upper Red	≤20		
Red	≤16	4) CRHT intervention required	9-20
Amber	≤13		
Green	≤8	5) No or low levels of CRHT intervention	0-12
Discharge	≤7		

Table 7.2 Indicating the scoring categories for Model 2. These categories were determined based on the information provided by the Receiver Operator Curve analysis and indicate overlap between the categories.

Figure 7.2 CRAFT Crisis Scoring Spectrum

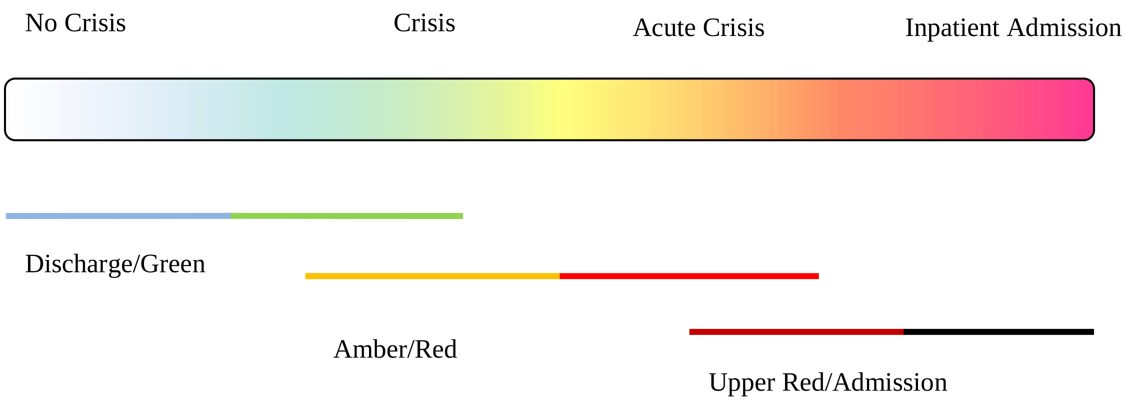
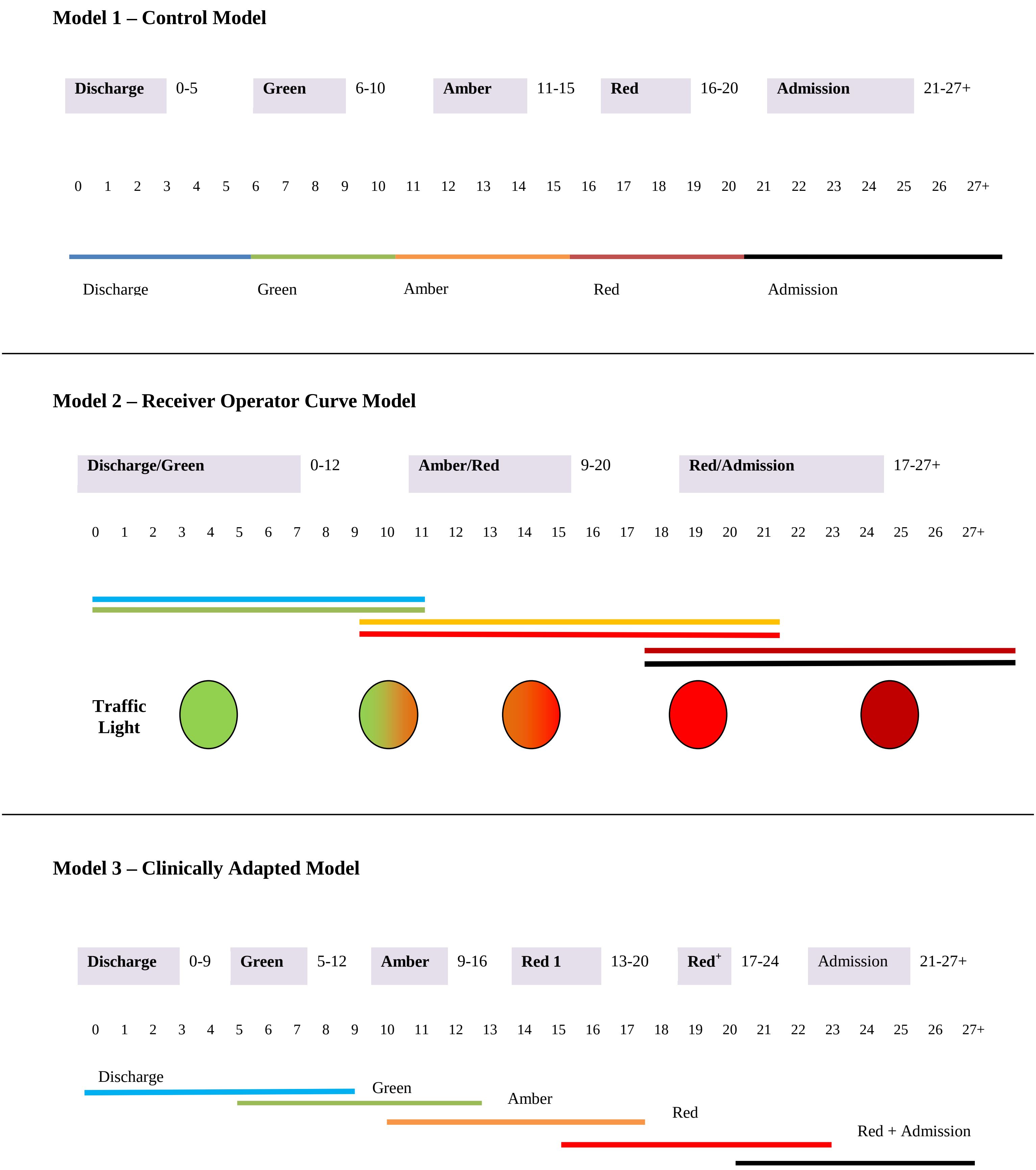


Figure 7.2 Outlining the CRAFT crisis scoring spectrum representing three main tiers of treatment, with areas of overlap conceptualised as periods of treatment transition.

The third model was developed as a more considered attempt for a model based on observation to provide a model that was clinically relevant and indicated areas of category independence as well as overlap. Therefore it would reflect the clinical reality as observed in practice and agreed as a suitable model with both the CRHT teams and the research team. Due to the clinical nature of the third model, the Red category was divided into lower and higher levels of input with visits once a day as the lower category and more than one visit a day in the higher category, which reflects the reality of clinical practice.

Figure 7.3 CRAFT Cut-off Models



7.5 Model Comparison – Sensitivity and Specificity Analysis

In order to demonstrate and compare the clinical utility of the models developed (outlined in Figure 7.3) they were compared for levels of sensitivity and specificity. Two comparisons were made between the models. The first comparison looked at the crude cut-off to differentiate between the treatment required group (Amber, Red and admission categories) and the no-treatment group (green, discharge and inappropriate referral categories). The second comparison looked at the ability of the model to accurately identify the level of treatment offered by the team.

In the pilot stage of the study, the assessor was asked to indicate the treatment colour status the patient had been placed on by the team. The status indicated was then compared to the outcome of the measure based on the different models (Figure 7.3). This is a method of assessing the concurrent validity of the measure by comparing the actual outcome to the measure indicated outcome (Anastasi, 1976). Where high levels of agreement are indicated between the actual treatment level given and the measure indicated status, the concurrent validity is high.

The main concern is to ensure that the crisis measure accurately identifies when patients require CRHT support and when CRHT support is not required. This is to ensure that individuals requiring support are accepted by the team for treatment and that resources are directed to those who need them most. If patients requiring CRHT support are not identified through the assessment procedures, they may not access the services appropriate for helping them manage their risks and challenges in order to make positive progress toward a helpful crisis resolution. A possible worst case scenario would result in the individual potentially harming themselves or others as a consequence of not receiving treatment when it was needed.

The levels of treatment indicated by the three scoring models were compared to the status indicated by the CRHT team at the time of scale completion. The number of True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) (as described in section 7.6.1 and 7.6.2) were compared between models to identify the most appropriate scoring model. Ideally, the outcomes of the measure are assessed against a gold standard such as a previously developed measure or a diagnostic criterion (such as the DSM IV), but these do not currently exist for acute mental health crisis.

Assessing the sensitivity and specificity of the scoring models will give an indication as to how sensitively and accurately the measure would be able to identify those requiring CRHT intervention and those for whom this intervention would be inappropriate should it adopt these scoring models.

If the scoring model is able to make this crude but important differentiation between those requiring CRHT support and those that do not, the second aim is to understand if the scale outcomes are able to indicate the level of CRHT input (or admission) required.

7.5.1 Comparison 1

The first comparison looked at 'CRHT intervention required' compared to 'CRHT intervention not required'. The most important criterion for the scale is the ability to indicate when a patient is experiencing crisis and requires input from the team and when a patient does not require input from the CRHT team. In particular, it is important for the team to identify patients requiring support so that they do not slip through the CRHT net.

The Green category is generally used by the team as a period of monitoring to ensure that the improvements observed are maintained with minimal CRHT input. If a patient is assessed and viewed to be 'Green' status, it is very unlikely that they will be taken on by the team as they could be monitored by other services such as their General Practitioner or Community Mental Health Team. From the analysis, it was identified in the frequency of the categories (Chapter 3, section 3.3, Figure3.2) that it was appropriate to collapse the three categories of no-crisis, discharge and green status together. Based on this understanding, where the Green category is stated it can be assumed that this individual is ready for discharge and therefore this was viewed as a non-CRHT treatment patient.

This analysis identified the number of True Positives, True Negatives, False Positives and False Negatives for the crisis measure in terms of identifying when a person requires CRHT intervention

(Amber, Red, Admission treatment categories) compared to when they do not (Green, inappropriate referrals and discharge categories). Below are the definitions of these categories in the context of this analysis:

- True Positive (TP) – the patient requires input from the CRHT as indicated by the team with Amber, Red or Admission status and reflected in the outcomes of the measure.
- True Negative (TN) – the patient does not require input from the CRHT as indicated by a green or discharge status and reflected in the outcomes of the measure.
- False positive (FP) – CRHT treatment was indicated by the crisis measure (Amber, Red or Admission) but not indicated as necessary by the CRHT team.
- False negative (FN) – support was not indicated by the crisis measure (Green or Discharge indicated) but treatment was indicated by the CRHT team.

The model comparison indicated if a TP, TN, FP or FN had been achieved. Frequencies and percentages of the TP, TN, FP, and FN categories were then identified and entered into online software to identify the levels of sensitivity and specificity using the clinical calculator (Lowry, 2011) to calculate the probabilities.

As indicated in Table 7.3, models 2 (the ROC developed model) and model 3(the clinically developed model) are shown to be the strongest models for differentiating between patients who had been shown to require CRHT intervention input and those who did not. Both models provided the same crude cut-offs for CRHT intervention required compared to CRHT intervention not required. These two models gave accurate indications of treatment need when compared to the treatment decisions made by the CRHT at the time of the CRAFT measure completion for 84.8% of cases when compared to the CRHT indicated status. In addition to achieving high accuracy for the identification of CRHT intervention to non-intervention groups, is to ensure that false treatment identification is kept to a minimum. Models 2 and 3 presented 8.1% of cases as indicating CRHT support that were not identified by the CRHT as requiring support (probability of 0.11) and therefore represents a slightly higher sensitivity compared to that indicated by the CRHT. Scoring models 2 and 3 identified

7.1% (probability of 0.27) of the sample as not meeting the criteria for CRHT support when they had been identified by the CRHT team as requiring support. Scoring models 2 and 3 achieved the strongest results in all four areas of True Positive, True Negative, False Positive and False Negative when compared to model 1. However, in some areas model 1 performed not too dissimilarly from models 2 and 3. For example, the probability of a True Positive for Model 1 was 0.84 compared to a probability of 0.89 for models 2 and 3. These outcomes compare well to more established measures such as the Hospital Anxiety and Depression Scale that has been shown to have sensitivity 0.89 and specificity 0.75 when used in the community setting (Olsson, Mykletun & Dahl, 2005).

Table 7.3 – CRAFT Scoring Model Comparison 1

	Model 1 Control Model				**Model 2** ROC Model				**Model 3** Clinical Model			
	Frequency	Valid Percent	Cuml Percent	Prob.	Frequency	Valid Percent	Cuml Percent	Prob.	Frequency	Valid Percent	Cuml Percent	Prob.
True Positive	177	59.8	59.8	.84	193	65.2	65.2	.89	193	65.2	65.2	0.89
True Negative	48	16.2	76	.56	58	19.6	84.8	.73	58	19.6	84.8	.73
False Positive	33	11.1	87.2	.16	24	8.1	92.9	.11	24	8.1	92.9	0.11
False Negative	38	12.8	100	.44	21	7.1	100	.27	21	7.1	100	.27
Totals	296	100			296	100			296	100		

Table 7.3 comparing the three developed scoring models for ability to differentiate between CRHT treatment required compared to CRHT treatment not required as compared to treatment decisions given at the time of completion of the CRAFT.

7.5.2 Comparison 2

The second set of comparisons compared the models for how well the scoring model indicated treatment levels appropriately (discharge – admission) in terms of reflecting the assessment outcome levels indicated by the assessor. The total score outcome was compared to the actual status of the patient's treatment as recorded at the time of assessment. The outcomes of the measure were shown to fall into one of the following four categories whose definitions differ in nature from the first analysis:

- True Positive (TP) – this shows that the level of crisis indicated by the scale outcome is the same as the crisis status indicated by the assessor.
- True Negative (TN) – this indicates that CRHT input was not required by the patient. Where the measure and CRHT indicate discharge appropriately (Green and Discharge status) a True Negative is present.
- False positive (FP) – is indicated when the scale outcome indicates a status level higher than that actually observed. This could mean a higher status level of crisis being indicated by the measure compared to the CRHT indicated level. For example where a Red status is indicated by the measure when Amber was indicated by the team.
- False negative (FN) – is shown when the scale outcome is at a lower level than that indicated by the team. This could mean crisis not being indicated when the crisis state is indicated by the CRHT or whereby the level of crisis indicated by the measure is at a lower level than that indicated by the team, e.g. Green being indicated by the measure when the CRHT indicated Amber.

It is indicated by the outcomes outlined in Table 7.4 that model 2 (the ROC developed model) was most able to accurately identify the level of treatment required (True Positive and True Negative) and was less likely to falsely identify higher or lower levels of treatment than needed (False Positive and False Negative).

The Control Model (Model 1) was the least accurate model as indicated by the lowest cumulative per cent by 19.6% for True Positive and True Negative and the highest valid per cent for False Positive (31.8%) and False Negative (30.4%), which was also reflected in the probabilities. Therefore although these models performed similarly for Comparison 1 looking at the crude cut off of CRHT treatment required compared to no CRHT treatment required, the accuracy of model 1 in terms of treatment level which is noticeably less than both models 2 and 3.

The ROC model (Model 2) was clearly shown to be the most accurate model, accurately identifying the treatment level 77% of the time with a probability of 0.80 for a True Positive and a probability of 0.69 for obtaining a True Negative. In addition, model 2 was less likely to identify a false positive (prob 0.20) or a false negative (prob. 0.31).

Table 7.4 – CRAFT Scoring Model Comparison 2

	Model 1 Control Model				**Model 2** ROC Model				**Model 3** Clinical Model			
	Freq.	Valid Percent	Cumul. Percent	Probability	Freq.	Valid Percent	Cumulative Percent	Probability	Freq.	Valid Percent	Cumul. Percent	Probability
True Positive	94	31.8	31.8	0.5	172	58.1	**58.1**	**0.80**	124	41.9	41.9	0.69
True Negative	18	6.1	37.8	0.16	56	18.9	**77.0**	**0.69**	46	15.5	57.4	0.39
False Positive	94	31.8	69.0	0.5	43	**14.5**	91.6	**0.20**	54	18.2	75.7	0.30
False Negative	90	30.4	100	0.83	25	**8.4**	100	**0.31**	72	24.3	100	0.61
Totals	296	100			296	100			296	100		

Table 7.4: CRHT intervention level indicated by the team assessment compared to the crisis measure indicated treatment level.

As mentioned previously, attempting to measure a latent construct as complex as acute mental health crisis is challenging. As a result of the construct of interest being removed from direct measurement, it has to be measured through items that are believed to represent some part of the underlying construct. This results in error between the observed level on the construct and the ‘true score’ or actual level on the construct. Therefore, it is expected that there will be some discrepancy between the observed score and the true score, which is often demonstrated when tests of specificity and sensitivity are completed. The aim in psychometric development is to provide a measure that shows ‘good enough’ levels of sensitivity and specificity. That is, the measure is both accurate most of the time and inaccurate as little of the time as felt clinically acceptable. Clinically, it is far more concerning and risky if individuals experiencing crisis and requiring support are not picked up through assessment. In cases where an individual is taken on by the team for treatment when intervention is not required, this is less risky in terms of the safety of the individual, but this could become draining in terms of resources for the team. This could lead to the CRHT team picking up a number of individuals who do not require their support. As outlined in crisis theory (Chapter 1), poorly resolved crises can result in a reduction in resilience and an increased vulnerability to relapse. Therefore, the focus is on keeping the number of False Negatives to an absolute minimum whilst ensuring that the number of True Positives is as high as possible. In this context, Model 2 appears to be functioning really well by having a high proportion of True Positives and True Negatives whilst having only a small probability of indicating either False Positives or False Negatives. When this is framed with the assumption that there will be some error due to the complexity and nature of crisis as well as error due to measurement error, it is clear that model 2 provides a useful scoring model for crisis assessment.

7.6 Summary and Conclusions

The experience, skills and knowledge of the multidisciplinary CRHT team are a vital resource from which significant treatment decisions are made for individuals experiencing an acute crisis. The aim of developing a psychometric outcome measure is to support the decision making process rather than replace it. Therefore the aim of developing cut-off points for the global overall CRAFT measure is to provide clinically credible and relevant indicators of need to guide and support treatment decisions in practice.

Based on the analysis of the sensitivity and specificity of the models outlined in sections 7.5.1 and 7.5.2, model 2 which was developed from the ROC analysis, was chosen.

The first model, where the scale cut-offs were crudely identified from dividing the scale into 5 separate categories, was shown to be fairly sensitive to crisis but at the expense of accuracy in terms of its ability to accurately identify treatment level. Model 3 was developed to try and reflect a more clinically relevant attempt at developing a model whereby treatment categories were more specific and overlapped. This was a comparative model which offered more differentiation between the treatment levels than the ROC developed model (Model 2) but with the acknowledgement of overlap between the treatment levels, which reflects the clinical observations from practice. This model performed equally with model 2 for differentiating between 'CRHT treatment required' and 'no CRHT treatment required' (comparison 1) as it had the same cut-off points but performed at a considerably lower level when trying to differentiate between the different CRHT treatment levels.

The ROC model (Model 2) was clearly shown to be the most appropriate model for clinical practice as a result of these analyses. The model provides broad cut offs that are able to guide mental health professionals toward 1) No or low levels of CRHT intervention required (Discharge category + Green category), 2) CRHT intervention required (Amber Category + Red category), and 3) Inpatient admission or intensive CRHT intervention required (Red category + Admission category). These broad overall cut-offs provide guidance and offer flexibility for individual need and reflect the current clinical preference for a continuum model rather than distinct categories differentiating between 'functional' and 'dysfunctional' individuals. The overall outcome of the scale acts as a guide for treatment planning with the ability to focus in on areas of strength and weakness using the outcomes of the individual subscales and where necessary individual items. This allows for a working partnership between the scale outcomes and the clinical expertise provided by the CRHT multidisciplinary teams.

However, the flexibility provided by the scale may also be viewed as a limitation to the model. The cut-offs are broad and encapsulate two possible treatment levels each, therefore making them more vague than those outlined by the other models. It only has three categories, which basically indicate low, medium and high CRHT treatment levels. Although useful, this does not directly reflect the treatment levels used in practice.

The Clinical Model (Model 3), which offers a 5 level model, showed some strengths in its ability to differentiate between treatment levels. Although this is not at the level of sensitivity and specificity required at this stage, it may indicate the potential for further development in the future.

Both models 2 (ROC model) and model 3 (Clinical model) present treatment levels that show overlap, which are more reflective of the treatment spectrum observed in practice. From observation and the statistical indication of previous analysis in this research (Chapter 6), the crisis journey appears to function along a spectrum. Patients move along this treatment spectrum or continuum, moving through transition periods from one crisis level to the next. The distinct treatment levels depicted in model 1 are less likely to represent an overview of the crisis journey, with the crisis levels appearing less dynamic and interactive than models 2 and 3. The crisis journey is rarely smooth or predictable, often comprising of steps up and down the crisis/treatment spectrum at different stages for different individuals. Where one person may take a more continuous and predictable journey, another's may be more disjointed with times of high activity experienced as rapid recovery or deterioration and at other times of much slower paced change. Therefore where one individual may be ready to move onto the next stage of treatment, another may require a longer period of watchful stabilisation with further progress to try and ensure recovery through that treatment phase before moving them forwards.

There were three main areas of concern for this measure. The first concern was the ability to capture the defining essence of a construct as complex as acute mental health crisis in order to identify potential items to contribute to the item pool. The second concern was that the 66 item dataset used for these analyses was taken from the original 143 item measure. This is quite a lengthy assessment tool which may have the potential for causing bias, fatigue or simply lack of engagement with completing the tool due to the demands it would place on mental health professionals' time. Finally, the third concern centred on the complexity of the 11 point item scale, therefore the potential for variance. Overall, it was an ambitious goal to develop a measure that would accurately measure crisis, within the individual patient context, that could offer treatment guidance with confidence. Potentially the outcomes of this research may not have been able to identify any meaningful patterns from which to develop a useful outcome measure. It is encouraging that model 2 (ROC model) shows good levels of accuracy overall. Future research should centre on obtaining a new dataset based on the 66 item measure to support the development and refinement of cut-offs further.

Chapter 8

Summary of the Validity and Reliability of the CRAFT

8.1 Introduction

This chapter will bring together all the different aspects of validity and reliability results relevant to the final CRAFT that have been presented throughout the preceding chapters, as well as reporting additional analyses relating to the temporal and inter-rater reliability.

Evidence of the validity and reliability of the Crisis Risk and Adaptive Functioning Tool (CRAFT) will provide it with essential quality indicators. The aim of measurement design is to deliver an approach that will draw out information to accurately represent the level on the variable of interest (validity) and to understand how consistent and stable the outcomes provided are (reliability). This is particularly important to the development of the CRAFT, which aims to enhance the pre-existing CRHT assessment and measurement techniques and ultimately to improve the information upon which critical treatment decisions are made. The validity and reliability of the CRAFT will specify how representative, trustworthy and stable the outcomes are for representing the constructs of the subscales and the underpinning dimension of *adaptive community functioning.*

Validity is accepted in this research as the primary consideration of measurement development on the assumption that this will lay the foundations for the development of good reliability. Therefore, the validity of this measure was critical from the very first stages of its design. As described in Chapter 1 (section 1.2.7.2), the model developed by Simms & Watson (2007) for substantive, structural and external validity was used as a framework upon which to hang the design of this crisis measure. Following the development of the measure, the reliability for internal, inter-rater and temporal reliability were assessed using statistical techniques from CTT and modern measurement techniques embedded in Rasch analysis.

This section of the research is divided into two parts. The first part gives a summary overview of steps taken to ensure the validity of the CRAFT. The second part focuses on the reliability and provides evidence indicating the internal stability, stability over time (temporal) and inter-rater consistency.

The outcomes of these sections will provide evidence for the quality of the measure and therefore indicate its accuracy, usefulness and ability to provide meaningful outcomes in practice.

8.2 Validity

The aims of this section are to summarise and provide evidence for construct validity (Chapter 2) and criterion validity (Chapter 7). The methodology for achieving these types of validity is fully outlined in the previous chapters (Chapters 2 and 7) and so will not be outlined again here but the main evidence summarised and brought together in a more succinct manner.

Good validity ensures that the construct of interest is the construct being measured. Interestingly, when Messick (1995) (Chapter 1, section 1.2.7.2) describes validity, the reference is to making judgements of validity, not absolutes. Therefore the validity of a scale is the provision of evidence to suggest the reasonableness of assuming that the scale outcomes represent a point along the spectrum of the underlying construct.

The emphasis for construct validity is on ensuring that measurement construction has its roots firmly placed in solid construct theory foundations. This is an on-going process for which there are a number of different approaches. A number of influential authors in the field of validity lean towards construct validity as the only necessary form of validity for scale development (Linn 1980; Loevinger, 1957; Messick 1989). Content and criterion validity are believed to simply form a part of this and it was the approach adopted for developing the CRAFT.

Construct validity for the development of the CRAFT was assessed against Simms and Watson's (2007) validity flowchart (Chapter 1, Figure 1.4). This model was adopted due to its clear step by step outline that provides a comprehensive overview of the process. The flowchart was changed and adapted to the needs of this research to clearly indicate phases where different methods of providing evidence for content, construct and criterion-related validity took place (Figure 8.0). The convergent and discriminant validity terminology previously used for this flowchart has been changed to predictive and concurrent validity in line with the methods of analysis used in this research (highlighted in yellow, Figure 8.0). The use of the Rasch model in itself assures validity due to the expectations and assumptions of the model. The flowchart in Figure 8.0 is used in this chapter as a guide to support the evidencing of the CRAFT's validity.

Figure 8.0 – Construct Validity Flowchart

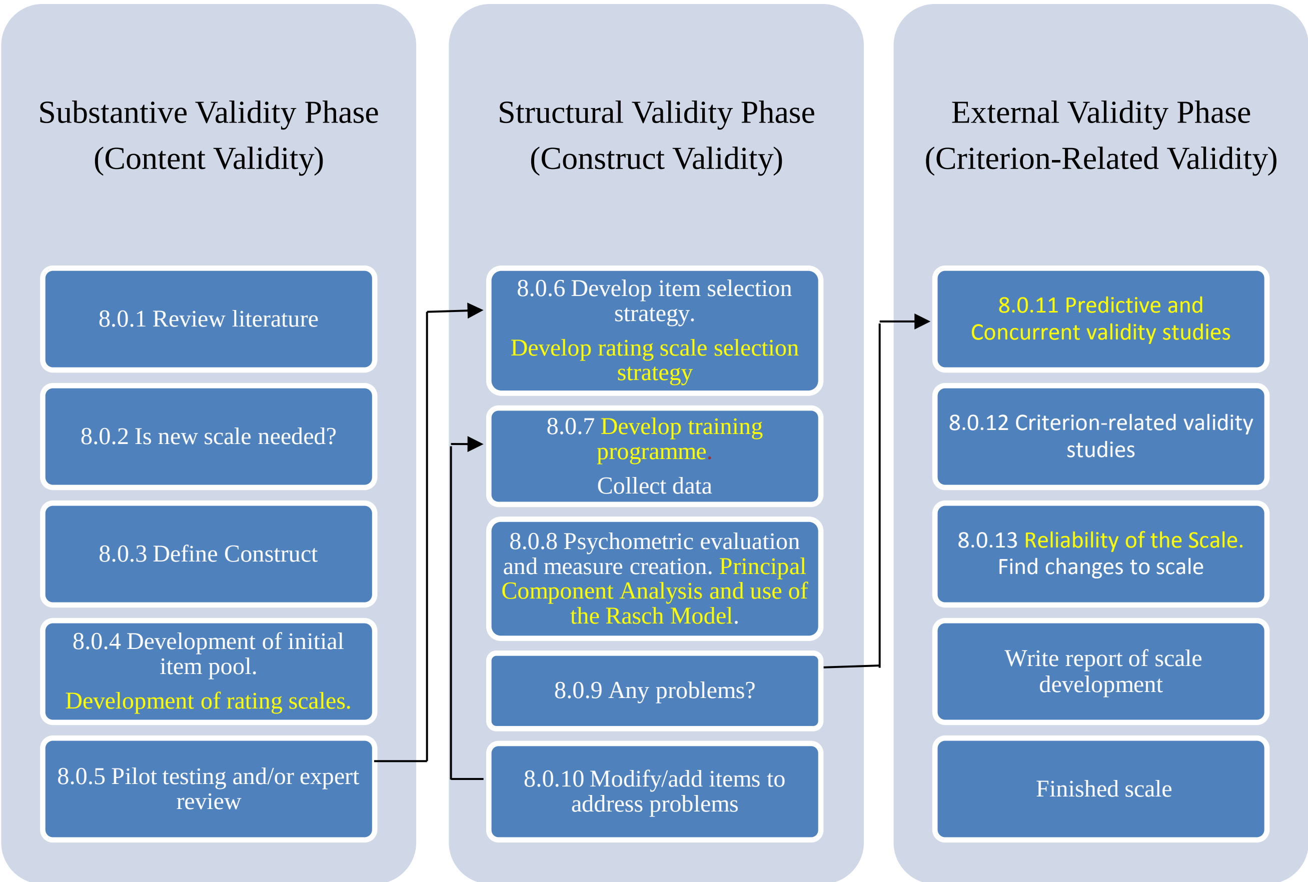


Figure 8.0 outlining the Simms and Watson (2007) construct validity adopted and adapted for the development of the Crisis Risk and Adaptive Functioning Tool (CRAFT). The text highlighted in yellow outlines the additional steps or analyses undertaken as part of this research for the development of the CRAFT.

8.2.1 Substantive Validity

The evidence for substantive validity will be drawn together and summarised here (Figure 8.0, sections 8.0.1-8.0.5). As outlined in Chapter 1 a comprehensive review of the literature was completed and indicated a significant lack of valid and reliable measures (Figure 8.0, section 8.0.1-8.0.2). The main attempt to develop an appropriate measure by Bonynge & Thurber (2008) had poor construct validity which did not meet the standards expected for modern measurement and failed to support comprehensive saturation of the construct and the development of a clinically credible rating scale. The literature indicated a continuing need for the development of a comprehensive, standardised, valid and reliable psychometric measure for the assessment of crisis by CRHTs. In terms of defining the construct (Figure 8.0, section 8.0.3 & 8.0.4), this research demonstrated a comprehensive approach to identifying the item pool and the item rating-scale. Content validity was evidenced in Chapter 2 through the comprehensive investigation of crisis assessment and the experience of crisis. This was achieved using qualitative techniques to extract a representative item pool which provided a sample of all the possible items that could be included. This is of particular importance in measurement design as item omissions at this stage of measurement development could result in a measure that does not fully represent the construct of interest or worse, does not represent the construct at all.

The next and final step in substantive validity is to evidence appropriate initial pilot testing to identify any teething problems or obvious changes (Figure 8.0, section 8.0.5). The initial outline of the CRAFT was shared with mental health professionals for feedback. The mental health professionals feedback was incorporated into the scale and the scale was then shared with staff and patient focus groups to refine and finalise the pilot scale (Chapter 2). Items were reviewed by the groups for relevance to crisis assessment and item use frequency as part of routine assessment. This is accepted to be an appropriate and thorough approach for the initial pilot and provides evidence of validity to support the first version of the CRAFT.

8.2.2 Structural Validity Phase

The next section on the Simms & Watson flowchart relates to the structural validity of the CRAFT. The item and rating scale selection strategy (Chapter 2) ran over two phases for the development of the first version of the measure. The first phase identified the initial item pool and rating scale from staff and patient interviews, which were then refined in the second phase through expert review by the research team and both patient and staff participant focus groups. This was a comprehensive approach towards the development of the first version of the CRAFT that aimed to ensure that both the practical and theoretical expectations of measurement development were met. Once the first pilot version of the item pool and item rating-scale had been chosen, a training programme was developed and implemented. It was decided to use a flexible approach to training and implementation to allow the training programme to develop and evolve in order to meet the needs of the teams. The final training approach involved small group training sessions followed by a one to one shadowing session with the trainer. Feedback from both those trained and from the trainers indicated that allowing the training to include a real world application of the measure was preferred and supported understanding. Both stages of structural validity are essential foundations upon which later reliability outcomes will stand. The comprehensive approach to item and rating scale selection stands the CRAFT on good ground to achieve sound internal stability (evidenced by Cronbach's alpha statistics, section 8.3.1). This more flexible approach to training may have led to inconsistencies in the inter-rater reliability which is shown in the outcomes of the statistical analyses (section 8.3.3). Although this approach resulted in weak initial outcomes for inter-rater reliability it is believed that the final approach to training, when applied across the entire sample, would result in strong inter-rater reliability. This could be shown with further data collection and analysis.

The crisis measure was piloted across the two CRHT teams and 385 completed scales collected. Using this data, the item pool and item rating scales were modified based on the outcomes of statistical analyses embedded in both CTT and Rasch Analysis (Chapters 3 and 4). The approaches used for this section of the validity model (Figure 8.0; section 8.0.8) were comprehensive and robust and provide sound evidence that the CRAFT has good structural validity.

8.2.3 External Validity Phase

Predictive, concurrent and criterion validity (Figure 8.0; sections 8.0.11 and 8.0.12) are subcategories of construct validity. It is not within the scope of this research to collect data to support the assessment of the CRAFT's predictive validity qualities. Understanding the predictive qualities of the CRAFT would support the CRHT teams to utilise the CRAFT for longer term care planning and to identify the need for, and shaping of, services . This data can be collected as part of a future research project based on the final version of the CRAFT.

Criterion validity would be assessed by comparing the outcomes of the CRAFT against the outcomes of a gold standard measure of the same construct. However, there are currently no valid and reliable measures that tap into the same crisis construct as the CRAFT and there is debate around the usefulness of comparing newly developed measures against previous measures when the aim is to develop and improve previous attempts, not simply replicate previous efforts. Therefore, even if there had been a valid and reliable pre-existing measure the usefulness of this approach to validity is questionable.

Concurrent validity data was collected by asking mental health professionals to indicate the status level (Red, Amber, and Green for example) of the patient at the time of scale completion. The scoring model developed was then compared against the actual outcomes decided by the team, looking at the sensitivity and specificity, e.g. the percentage of accurate results against the number of false positives and false negative outcomes. Overall, the results were encouraging and suggest that the scale is able to reflect and support the treatment decisions of the team (Chapter 7).

8.3 Reliability

Tests of reliability check that measurement outcomes are dependable and consistent and form one of the final steps on the Simms & Watson (2007) construct validity model (Figure 8.0, section 8.0.13). Statistical methods in CTT and Rasch analysis aim to understand how well the observed score represents the true score (Suen, 1990). The higher the reliability coefficient, the closer the observed score is to a linear relationship with the true score. As mentioned in Chapter 1 (section 1.3.3.2), there are three main types of reliability that a researcher is interested in when developing a measure and these are assessed here; 1. Internal consistency (8.3.1), 2. Test-retest (8.3.2) and 3. Inter-rater reliability (8.3.3).

Reliability is assessed by studying the relatedness between two sets of scores. For internal consistency this is achieved by calculating Cronbach's coefficient alpha based on the item and sum scale variance, for temporal reliability this is a correlation to look at the relationship between time one and time two of scale completion and for inter-rater reliability it is the correlation between rater one and rater two.

8.3.1 Internal Consistency

Internal consistency tests that the items within the measure have the same value of the latent construct. If the measure meets the criteria for internal consistency then the subtotals of the parts should indicate the same level of the latent construct when the measure is divided in half.

Cronbach's alpha is a measure of internal stability of a psychometric scale. Developed by Cronbach (1951), it is a reliability coefficient that represents how stable a measure is ranging from 0-1 with 0 being completely unstable and 1 being perfectly stable. It is the aim of measurement development to create a scale that is as stable as possible so that the measure and its parameters do not change and therefore the outcomes are consistent and meaningful. Nunnally (1978, p.245) suggested that a coefficient alpha of ≥ 0.7 was good but also suggested that increasing it much beyond 0.8 would not be that advantageous. However, in the context of a scale that may be supporting mental health professionals to make decisions about accepting potentially very risky patients for treatment, it would be considered an advantage to increase the reliability of the scale to as high as possible. George and Mallery (2003) provide the following rules of thumb: "Excellent > 0.9, Good> 0.8, Acceptable >0 .7, Questionable > 0.6, Poor > 0.5, and Unacceptable <0 .5" (p. 231). To calculate the coefficient alpha for internal consistency requires a large sample to compensate for any inadequacies in the data due to sampling error. Violation of the underlying assumptions can result in over- or underestimation of reliability.

Internal consistency was tested using Cronbach's alpha. Cronbach's alpha calculations were carried out on the transformed item scales following Rasch analysis. Cronbach's alpha was shown to be excellent at α=0.972 (N=310). Cronbach's α was not shown to improve through deletion of any of the items.

8.3.2. Test-Retest/Temporal Reliability

Temporal reliability gives an estimate of a measure's stability over time. The principle here is that where an individual's presentation of a construct is stable, the measure should give similarly stable, consistent outcomes over time. Therefore to test temporal reliability the measure is administered at two time points to the same person and the outcomes compared. The time period between administering the measure can be anything between a few hours up to several years but, it is more likely that reliability will decrease with longer time intervals as the opportunities for change become greater. Therefore it is important to use a time period that is long enough to demonstrate stability but not too long that natural change would influence the outcomes.

The underlying assumption for temporal reliability is that where the individual being measured has not changed over time, the latent construct has remained stable. Therefore, it is particularly difficult to obtain data for temporal reliability from a crisis population due to the changeable nature of the presentation. Poor temporal reliability outcomes would most likely represent the stability of the sample rather than the stability of the measure. In this context, it is more appropriate to obtain data from a different but comparable population such as an inpatient population.

Measurement error may occur due to changes in the assessor, the assessed or the environment in which the assessment takes place. For example if the assessor is fatigued, the assessed is physically unwell or on the day there is a power cut and the heating and lighting is not working in the building, could all have an impact on the reliability outcomes. One of the main concerns with collecting data for this form of reliability is error due to practice effects when the participant completes the measure and then has time before the next administration to practice that skill area to improve their outcome. For the purposes of the CRAFT the patient themselves does not complete the scale but the mental health professional completes it based on observation. Therefore, theoretically, practice should improve the accuracy of the outcomes.

8.3.2.1 Appraisal of Temporal Reliability Sample

The sample used to obtain data for temporal reliability for this measure was from a comparable but stable population. As mentioned above, the crisis presentation is marked by instability and rapid change, and therefore it would be difficult, if not impossible, to indicate the stability of this measure based on data collected from the crisis population. Therefore, data was collected from an equally acute and risky population but residing as inpatients at a Low Secure Mental Health Unit in Luton called the Robin Pinto Unit. This unit has approximately 14 residents at any one time who present with acute mental health difficulties that cannot be managed independently in the community. Completed crisis measures were collected for all residents in August 2010 and July 2011 at two time points approximately 2 weeks apart. Data collection was carried out twice to obtain enough data to reach reasonable power as indicated by Cohen's power calculation (1992) which indicates that N=22 is required to achieve a large effect size at 0.1 (p.158). Due to the stable nature of the Robin Pinto Unit residents it would be expected that little change would be shown between the two sets of data collected.

8.3.2.2 Total Scale Score Temporal Reliability

The temporal reliability was calculated using Spearman's *r* for non-parametric data because the data only met one of the two assumptions. The data met the assumption for interval scaling due to the Rasch transformations but did not meet the assumption for normally distributed data, being skewed to the lower bound. There was a significant relationship between the two sets of ratings, (N=23), *r* =0.971, *p* (one-tailed) <0.001. The percentage of variability shared is the percentage of the variation in one variable that is shared by the second variable. This is calculated as r^2 x 100 and therefore was calculated as 94.3% of the variability is shared.

8.3.2.3 Subscale temporal reliability

The temporal reliability was calculated using Spearman's *r* (N=23) for non-parametric data (Table 8.0). The temporal reliability demonstrated by the scale was >0.7.

Table 8.0 Subscale Temporal Reliability Outcomes Table

Subscale	Spearman's r	Significance p	Percentage of variance shared
1. Crisis Recovery Indicators	.967	.001	93.5
2. Adaptive Decision Making	.954	.001	91.0
3. Risk of Harm to Self	.937	.001	87.8
4. Mediating Factors	.979	.001	95.8
5. Daily Structure	.835	.001	69.7
6. Risk of Harm to Others	.843	.001	71.1
7. Feelings and Affect	.956	.001	91.4
8. Basic Needs	.764	.001	58.4

Table 8.0: Table confirming that all of the subscales achieved acceptable to excellent levels of temporal reliability beyond chance levels. This outcome is further explained by the variance shared statistic.

8.3.3. Inter-Rater Reliability

Inter-rater reliability is a measure of how much consensus different raters have when using the same scale (e.g. Bliese, 2000; Lebreton, Burgess, Kaiser, Atchley, & James, 2003). Where inter-rater reliability (IRR) is poor suggests that either there is a problem with the scale itself or the method of completion is inconsistent. Lebreton and Senter (2008) define IRR as "IRR refers to the relative consistency in ratings provided by multiple judges of multiple targets". Inter-rater reliability is a method of checking if outcomes remain stable when completed by different assessors on the same subject. This is assessed when analysing objective assessment measures. By looking at the differences in measurement outcome from two different assessors on the same patient it is possible to examine how stable the measure is between different raters. The concept here is that where scoring instructions are clear the assessors will rate items in similar ways and therefore obtain similar outcomes. Where scoring instructions are ambiguous, there is a greater chance of the measure being completed differently by different assessors, therefore the outcome of the scale is different and consistency between raters will reduce.

There are a number of approaches to inter-rater reliability including Cohen's kappa (1960) and Fleiss's kappa (1971) for categorical data and Pearson's and Spearman's correlation coefficients for continuous/parametric and ordinal/nonparametric data. Nunnally (1978) suggests that a cut off of 0.7 is necessary for a measure to have good reliability and this cut off will be adopted for the purposes of this research.

8.3.3.1 Sample Size

Data was collected asking CRHT mental health professionals to complete two crisis measures, completed by different staff members on the same patients, after the same assessment and at the same time. This was essentially asking a second member of staff to repeat the crisis measure documentation for a patient. We were able to collect 18 pair sets of crisis measures under these strict criteria . This is 4 short of meeting Cohen's power for the lowest specification (N=22 for power 0.1 and large effect). It was not felt appropriate to persist with this exercise to achieve better data collection due to the pressured nature of the CRHT work and the pressure this may place on staff. However, it was also felt unethical to not include this data in the research based on the time taken by team members to support with this exercise. The limitations of the analysis are outlined in the next section (section 8.3.3.2).

8.3.3.2 Limitations of the Inter-Rater Reliability Analysis

The sample considerations outlined above combined with Trust concerns around the possible distraction of staff from key work tasks resulted in the decision that it would not be ethical to pursue the collection of large amounts of data simply for the purposes of obtaining the inter-rater reliability. It was also recognised that motivation to complete the measure is important to ensure validity as this relies on representative data. CRHT staff working in the highly pressured context of community crisis work could be hypothesised to have low motivation for completing the same measure twice on the same patient due to the time limitations they constantly battle. It is therefore less likely that this data will be representative of the true inter-rater reliability of the measure. It was felt unreasonable to ask the teams to effectively repeat a 143 item measure simply for the purposes of this research at this time. However, in its final 66 item format, with the aim of implementing it as part of standard routine practice, there may be more opportunity to collect this data as part of future research.. In addition, it is recognised that the training approach implemented significantly changed over the course of the research and therefore differences found between rater's scores may simply reflect the differences in training received rather than a fundamental difficulty of the measure. The lack of statistical power, the motivation of staff to complete the measure twice and the significant changes in training approach are significant limitations for calculating the inter-rater reliability. However, it was decided to use the data collected to calculate and report the inter-rater reliability whilst acknowledging that the outcomes are not likely to be valid.

8.3.3.3 Total scale Inter-Rater Reliability

The inter-rater reliability was calculated using Spearman's *r* for non-parametric data as the data only met one of the two required assumptions. As outlined above, the data met the assumption for interval scaling due to the Rasch transformations but did not meet the assumption for normally distributed data, being skewed to the lower bound. There was a significant relationship between the two sets of ratings, (N=18) $r = 0.698$, p(one-tailed) $<.001$. The percentage of variability shared was 48.7% for this sample. The outcome falls below the 0.7 cut-off, suggesting that the relationship is a weak one. However, this could be rounded up to 0.7 if rounded to one decimal place, which would bring the outcome to just within the acceptable parameters of inter-rater reliability.

8.3.3.4 Sub-scale Inter-Rater Reliability

Each of the sub-scales were individually analysed for inter-rater reliability. Where both parametric assumptions were not met, the non-parametric alternative of Spearman's *r* was calculated. The sample used to calculate the correlation was N=18, which is small and therefore provides only an indication of the inter-rater reliability. All analyses were one tailed. The results of the analyses are shown in Table 8.1 Spearman's *r*. Subscale 4 suggests that there was little if any relationship between the two rater's scores ($r<0.3$). Subscales 1, 5, 6, and 8 demonstrated a weak relationship between the two raters scores (r<0.7) and only subscales 2 and 3 demonstrated an acceptable level of relationship (r>0.7).

Table 8.1 Subscale Inter-Rater Reliability Outcomes Table

Subscale	Spearman's r	Significance p	**Percentage of variability shared**
1. Crisis Recovery Indicators	.562	.01	31.6%
2. Adaptive Decision Making	.780	.01	60.8%
3. Risk of Harm to Self	.780	.01	60.8%
4. Mediating Factors	.297	.01	8.8%
5. Daily Structure	.407	.01	16.6%
6. Risk of Harm to Others	.428	.01	18.3%
7. Feelings and Affect	.366	.01	13.4%
8. Basic Needs	.681	.01	46.4%

Table 8.1 Inter-rater reliability statistics were shown to be weak but it is recognised that there were a number of limitations that restrict the accuracy of these outcomes.

Overall, the inter-rater reliability was poor and remains questionable. However, this is considered to be a function of the preliminary scale which was considerably longer in length (more than double the size of the 66 CRAFT), the nature of the evolving training programme that continued to change throughout the data collection phase, the small sample size that did not meet Cohen's (1992) criteria and the difficulties obtaining data due to the pressurised nature of the CRHT work.

8.4 Rasch Indicated Reliability

The reliability of the crisis measure as indicated by the Rasch analysis is outlined in Chapter 4 (section 4.8, Table 4.1). Overall, the reliability indicated by this statistic reflected the findings from the CTT approaches outlined above, with good reliability for all of the subscales.

8.5 Summary

There are two main theoretical and statistical concerns in measurement development – the validity of the scale, e.g. is it measuring what we want it to measure, and the reliability, e.g. is the measure able to provide consistently meaningful outcomes?

8.5.1 Validity

There are no certainties in the judgement of validity. Theoretical methods and models of validity are a fundamental foundation upon which the development of a measure can rest, increasing the potential for a strong and stable measure once it is fully developed. It guides the process from the initial stages through to completed scale development. Where solid theoretical models are used as the foundation to measure development, validity is enhanced and sound reliability more likely to result.

The validity of the scale can be estimated and judged based on the method of construction utilised. The methods chosen for this research were based on a comprehensive model utilising the knowledge, skills and experience of the individuals for whom the measure ultimately aims to support - the patients and the mental health professionals of Crisis Resolution and Home Treatment. The theoretical model utilised was expanded to support theoretical underpinnings of Item Response Theory based in the Rasch model by focusing at both the item and scale level of measurement development. This was done by developing both the items and the item rating scales using comprehensive methods with the aim of achieving content saturation and interval level scaling, therefore developing a more valid and representative overall scale. The rating scale itself was particularly important because it was aiming to reflect the complexity of assessment, incorporating both risk and protective factor considerations as described by the participants in interviews. Agreement across both focus groups suggested that the final rating scale adopted has clinical strength and will support mental health professionals to rate items and to communicate the outcomes with patients in a manner that is meaningful to both. The patient participant group paid particular attention to the language of the scale and felt that the wording of *Cause for Concern* was clear and made good sense of how the scale items were assessed.

The Rasch model allowed focus at the item and scale level, supporting a move from the ordinal scale level to interval scaling. This strengthened the validity of the scale, ensuring that the scale used to represent items and subscale totals is at the interval level desired for accurate and true measurement.

Further to the construct validity demonstrated through the approach to scale construction, the concurrent validity of the scale was then shown in the results of the statistical analysis comparing the scale's outcomes to the actual CRHT team decision. This comparison provided evidence of the scale's sensitivity and specificity (Chapter 7), indicating that the scale is accurate both in terms of deciding between the crude cut off of crisis to no-crisis but also in terms of accurately estimating the level of treatment required. The next step for research would be to understand the predictive qualities of the scale, for example, the scale's ability to identify those most likely to recover from crisis and most likely to be a risk whilst experiencing crisis.

A possible limitation to the true validity of the scale is that the data was only collected from the Bedford and Luton CRHTs based on their patient populations. Therefore the scale may only reflect the needs of this population and may not be generalizable to a wider population.

8.5.2 Reliability

Overall, the reliability of the CRAFT and the subscales was shown to be acceptable to excellent for internal consistency and inter-rater reliability. Assessing the individual subscale reliability outcomes indicated that the reliability overall for the temporal reliability subscales was acceptable/good and significant. However, the inter-rater reliability demonstrated weaknesses.

The literature suggests that where poor inter-rater reliability is shown, a training need is indicated (Colton et al 1997). Wolfe, Koa & Ranney explain that "scorers with different levels of scoring do not focus on different (product or performance) features, but probably have different levels of understanding about the scoring criteria." (1998, p.465) The low inter-rater reliability outcomes for half of the subscales may reflect the training need identified after the first round of training, which resulted in a change from large to small group training and the subsequent addition of shadow sessions. Alternatively, the difficulties demonstrated may also reflect individual differences in crisis construct perception which would indicate a training need in terms of knowledge and understanding of mental health crisis in general. Both hypotheses relate to training gaps for specialist CRHT teams who currently do not receive any specific training. Without any specialised training to support knowledge, assessment and treatment of acute mental health crisis it is reasonable to assume that the conceptualisation of crisis between teams and between individuals may vary, which would ultimately result in differences in the way the crisis presentation is rated.

8.6 Conclusion

Overall, the CRAFT has demonstrated good levels of validity and reliability, based on a comprehensive and theoretically sound model for construct validity. The experientially rich data source of both crisis mental health professionals and patients, the statistically sound methods of scale refinement and the outcomes of the internal consistency and temporal reliability statistics demonstrated that the first version of the CRAFT will produce reliable and consistent results over time with further consideration for future training methods indicated. For the first pilot of a comprehensive scale, measuring a construct as complex as acute mental health crisis, the outcomes of this study are promising and further work toward improving the inter-rater reliability is realistic.

Chapter 9

Conclusions and Implications

This research was completed over a number of stages which have been reflected in the sequence of these chapters. Each chapter has offered an overview of a stage, an outline of what was achieved and a comprehensive summary from which to move forwards to the next chapter. To avoid too much repetition, this chapter offers a focus on the achievements, strengths and limitations of this project before going on to explore the possible clinical implications and utility of these findings and finally closing with a consideration of areas for future research.

9.1 Achievements and Strengths

The overall aim of this research was to develop an outcome measure that would support mental health professionals working in CRHT teams to make more accurate and meaningful treatment decisions within the clinical community context in which they work. The aim was to do this whilst ensuring both risk and protective factors were considered across the entirety of assessment and therefore the measure. This research has been committed to ensuring that the measure developed has sound construct validity and as a consequence to overcome the obstacles that have previously hindered measurement development attempts in this area. Completion of an in-depth investigation into the key characteristics of crisis, following on from recent work published in the literature, looked at the perspective of those who are experts in the field (patients and staff of the CRHTs) to produce a representative item pool. This has provided insights into the core themes that relate to the necessary considerations for crisis and home treatment assessment in the community context and for the development of a clinically credible item rating-scale. This was a collaborative achievement between the research team, the CRHT teams and experts by experience, the patients themselves.

One of the particular strengths of this research is the focus it gave to the development of the item rating-scale which had equal focus to that given to the development of the item pool. It was decided at the inception of the research that there had to be a commitment to the development of a clinically credible and valid item rating-scale to ensure that the information obtained on the items would provide meaningful and relevant information to clinicians and patients. It is an achievement of this research that this has been realised. This focus was maintained from the initial conceptualisation of the rating scale to encapsulate both risk and protective factors, throughout the process of rating scale refinement to ensure that each item was represented by a healthy functioning scale, to finally ensuring that the combination of the item information at the subscale level resulted in an interval level scale, a fundamental pre-requisite of parametric analysis. As a result of the attention given to the item rating-scale future quantitative research into crisis can now be completed using a measure that confidently represents interval level scaling at the subscale level and will therefore provide meaningful outcomes from parametric analysis. Finally, in relation to the item rating-scale, the decision to make the 'not applicable' category redundant was a clinically significant step forward, recognising that where 'not applicable' may be assigned would indicate a protective factor. In this particular area of measurement, 'not applicable' could in some circumstances provide misinformation by failing to capture a person's strengths. For example, where an individual is assessed as not having previously attempted suicide, this would be considered a protective factor by professionals working in mental health who recognise from the research that previous attempts are a significant predictor of future attempts and completion of suicide. In this scenario, the item 'Regret of previous suicide attempt' would be marked as '*not cause for concern'* on the CRAFT to represent the protective nature of this item response rather than rating this item as 'not applicable', which would fail to represent the protective nature of this outcome.

A further achievement of this research has been the application of both CTT and Rasch analysis in concert, recognising the strengths of both approaches and applying them in the development, refinement and evaluation of the CRAFT. Rasch analysis has popularly been applied as a confirmatory approach for assessing the structure of long standing measures such as the HADS and the Beck Depression Inventory. This research adopted Rasch analysis to assess the functioning of this measure from the very first stages of analysis at the item, subscale and whole measure levels and as a consequence, the first version of the CRAFT demonstrated evidence of good validity and reliability.

This research was carried out from a clinical perspective using modern statistical approaches to support the process of developing a clinical measure. The clinical perspective was paramount to ensure that the measure developed would provide clinical utility and support clinicians to complete crisis assessments. The statistical analyses suggested that the global overall measure was not unidimensional. This suggested that it would not be a legitimate step to total the item scores to provide an overall score on which to develop cut-off indicators of adaptive community functioning. However, there was evidence to suggest an underlying bi-factor model with a dominant underpinning core factor and therefore although it would not be a legitimate step to total the item scores, it was decided that it would still be helpful to understand the overall picture provided by the items information using an alternative approach. This research used the approach of combining the subscale percentile rankings to investigate cut-offs that would provide a meaningful pattern which could guide clinicians toward appropriate treatment decisions. The clinical utility of these cut-offs far outweighs any potential statistical drawbacks that may be argued. The main aim here is to support CRHT teams to deliver services in a more effective and safe manner. The CRAFT has clearly proven its ability to do this through the outcomes of analysis showing good levels of reliability, sensitivity and specificity. Therefore, implementation of the CRAFT as part of assessment by CRHT teams will enhance their appraisal of risk and protective factors to inform decisions regarding individual ability to adaptively function in the community, ultimately signalling the level of intervention required.

In summary, this research has produced the first clinically credible and statistically sound measurement tool (CRAFT, Appendix 18) to support crisis assessment and enable quantitative research to be completed in the acute mental health crisis population.

9.2 Limitations

The development of the item pool to uncover the construct of crisis was comprehensive and a leap forward in comparison to previous measurement development attempts. However, the qualitative aspects were not fully realised within the parameters of this research. Interviews were completed with staff and patients for the purposes of identifying the item pool (Chapter 2) using the basic approach of content analysis. However, it would have been interesting to fully investigate the experience of crisis using a more in depth approach such as Interpretative Phenomenological Analysis (IPA) and to have completed triangulation by interviewing carers to obtain their personal crisis perspective.

It is important at this point to recognise that the 66 item measure developed from the original 143 item pool has not been piloted in its own right. In addition, training needs identified in the pilot phase of this study were not implemented across the whole sample and therefore the combination of the shortened scale with the improved training approach should provide an improvement in measurement quality for future pilots from which the outcomes of these analyses could be reviewed.

The sample size was comfortably greater than 300 as suggested by Kass & Tinsely (1979). A larger sample would be preferable to achieve excellent power but this would be a leap to a sample of 1000. With the recognition that it took over 9 months of training and data collection to obtain the 385 measures used for the analysis, it is clear that this could not be achieved within the scope of this research. However, with the item pool reduced by more than half, it would now be more feasible to obtain this larger sample for future research.

Inter-rater reliability was shown to have poor outcomes in the analysis completed in this research, but as explained in Chapter 8, the training changed over the course of the research and the sample used was too small to provide confident conclusions. It would be expected that with the 66 item measure being piloted in its own right, with a more consistent and comprehensive approach to training and obtaining a larger sample, a more accurate assessment of the inter-rater reliability could be achieved.

Another possible limitation for this research was that the data was skewed to the 0 point for all of the items in the CRAFT measure. For the purposes of parametric analysis this could be viewed as a weakness. With the skew going towards the 0 point on the measure, this is more likely to downwardly bias any outcomes of parametric analysis, making it less likely that a Type 1 error would occur, i.e. significant findings would be shown when the null hypothesis is true. Therefore if outcomes are shown to be significant with a downward bias, this provides greater strength to the outcomes and reduces the likelihood that the outcomes occurred by chance. In addition, research has shown that parametric analyses are robust enough to manage skewed data (Gangestad & Thornhill, 1998). A normal distribution is not an assumption of Rasch analysis and therefore a decision was made to retain the original data rather than pre-process it as this would actually provide greater strengths in the findings than if the data had been processed prior to analysis. However, it is recognised that the statistical 'ideal' is to have perfectly normally distributed data, but clinically in the real world crisis population context, this would be very difficult to achieve.

A limitation for obtaining the data for temporal reliability is that it was obtained from an inpatient population rather than the crisis population. Although it would be preferable to obtain data from the population of interest, collecting data for the purposes of assessing the stability of the measure over time would not be possible using a crisis population. The literature suggests collecting data over two time points preferably a week to two weeks apart. A person experiencing crisis can show significant clinical change over the course of 24 hours let alone a week or two week period. As a consequence this would have resulted in evidence of poor temporal reliability which would not reflect the stability of the measure but the stability of the population. Therefore, although this was a limitation of the research it was one that was addressed using an appropriate alternative sample population.

In terms of the limitations of the training approaches used, the validity and reliability of the measure are directly affected by the expertise of the individual completing the measure. If there is not sufficient understanding of the techniques for rating, scoring and interpreting the scale, the validity and reliability of the scale will be automatically challenged. Therefore, a more comprehensive training approach would be recommended for future training purposes and inter-rater reliability to be reassessed on the basis of data collected from this sample. It is not unusual in mental health services to require comprehensive standardised training for the completion of complex measures, for example the HCR-20 for risk of violence requires all practitioners to complete a training programme and the measure itself should be completed by a multi-disciplinary team rather than by an individual. This assessment scale aims to accurately understand one of the most acute and complex mental health populations assessed and treated by mental health professionals. It is therefore not unreasonable to expect that a significant staff training initiative would be required to assure proficient use. One of the comments commonly expressed by staff participants was the benefits of the CRAFT training due to the educational element and direct application of this knowledge to practice. Verbal feedback received indicated that training and implementation of the scale as part of routine practice supported mental health professionals to be more vigilant to both risk and protective factors in the areas highlighted. Currently there is no specific CRHT training for mental health professionals entering NHS CRHT services. Therefore the CRAFT could potentially support CRHTs as a training tool as well as an assessment tool to support team treatment decision making.

It is acknowledged that this research presents a number of limitations and areas where the research could be improved. However, it is felt that this research has still provided a significant step forward for developing a useful and meaningful assessment tool for crisis that will support improved treatment decision making by mental health professionals working in this field. The limitations outlined here may provide useful areas for future research in this area and therefore support the continuing development of increasingly accurate crisis measurement tools.

9.3 Clinical Utility

The CRAFT was ultimately developed to act as a practical clinical tool and therefore its clinical utility is paramount to this research. The approach to the development of the CRAFT aimed to meet statistical, clinical, theoretical and practical considerations for the assessment of crisis. The measure has been shown to offer a valid and reliable approach to assessing an individual's ability to adaptively function in the community taking into consideration both risk and protective factors. It utilises a flexible scoring system that is able to offer the choice of a single or range rating system to encapsulate both stable and varying presentations on items. This flexibility has particular clinical credibility for a measure aiming to capture the essence of the crisis presentation for which changeability is a key feature. In addition, through the investigation of clinical, professional and personal experiences and the statistical techniques adopted to assess these variables, it is believed that the final item pool comprehensively represents the key features of crisis assessment in a measure that has been shown to take approximately 10 minutes to complete following a clinical assessment. This gives the CRAFT the potential for powerful clinical advantages due to its ability to capture the presentation of crisis in a concise but in-depth manner. The temporal reliability demonstrated for this measure provides the evidence required to show that it will be a useful tool for monitoring progress and the sensitivity and specificity of the cut-offs indicates it will also provide meaningful guidance in relation to treatment decisions. The good levels of temporal reliability support the use of the Reliable Change Index as an approach for assessing change in patients over time. The Reliable Change Index would provide a more clinically relevant and individually tailored approach for assessing change in crisis and therefore would support clinicians to meet individual need in a service that aims to adopt a person centred approach.

The identification of the 8 key characteristic areas of crisis assessment provides crucial guidance to mental health professionals working in crisis. The areas identified clearly outline the substance of a crisis presentation providing a useful guide to clinicians. The identification of these crisis characteristics guide clinicians towards an awareness of the individual's ability to cope and to be resourceful and to appreciate how stable and predictable a person's presentation is so that they can be treated in the community with confidence. This focus on the individual and their resources in terms of their ability to cope and self-manage links in well with the current shift in treatment models in mental health towards a recovery model based on empowerment, ownership, recovery and maintaining wellness. The statistical indications gave particular weight to a person's ability to think clearly and rationally in order to provide a strong foundation upon which to make decisions. It is not difficult to appreciate from a logical perspective that the ability to think clearly, to be able to make rational decisions, will be of particular importance for an individual being treated in their own home and a necessity for that person to function adaptively in the community. As expected, risk of harm to self and to others were identified as principal assessment areas and links in with the shift from inpatient to community treatment where risk assessment becomes a principal concern. Interestingly, mediating factors were identified as a key characteristic for crisis assessment, helping to highlight where social networks can be mobilised to provide additional support so that treatment can be delivered safely at home within a safe community structure. Basic needs and daily structure were two further subscales identified that link in with Maslow's hierarchy of needs. The clinical implication of these subscales indicates that where an individual is shown to struggle with attending to their most basic needs, they would need more practical support to establish this basic baseline of functioning. These more basic requirements would need to be in place before more in-depth therapeutic intervention could be delivered.

As outlined in the section above, the CRAFT requires an in-depth training programme similar to other measures aiming to measure presentations as complex as crisis. The training programme developed for this measure may be a step towards answering the problem of training for the specialist services provided by the CRHTs. At present there are no specialist training programmes for clinicians entering into this field of practice and therefore the guidance given through the CRAFT training may support mental health practitioners to understand both the basics of crisis, e.g. the 8 key areas of crisis assessment as well as more in-depth crisis characteristics such as those highlighted through the Rasch analysis. As part of this research a training programme and user manual were developed to support staff training and implementation of the CRAFT. This will be updated to reflect the final version of the CRAFT and can be used by teams to implement this assessment approach into CRHT teams providing specific guidance relating to the measure's unique features and providing evidence of the validity and reliability of the measure. There is also the potential for the CRAFT to be delivered through a basic and higher level training approach. The basic training would enable the clinician to complete the measure to provide basic outcomes relating to the level of treatment recommended for the patient. A higher level training would support clinicians to interpret the information provided by the CRAFT, for example, to be able to identify and interpret the information provided by the specific risk indicators and characteristic subscale items.

The CRAFT may also act as a helpful guide for patients in helping them to identify and understand their areas of strength and weakness, concisely summarising this complex mental health presentation. With a movement in mental health services towards recovery models, with a focus on the individual taking ownership for their own recovery through self-management, it is important that mental health practitioners support patients to do this by providing accurate information regarding their main areas of strength as well as difficulty or struggle. Identifying these key areas of challenge empowers the individual to address these difficulties for themselves where possible.

The development of computer software for this measure is currently underway. It will support the scoring aspect of this measure and will further improve its clinical utility. In addition, the possibility of handheld devices in the future could offer the potential for this software to be taken out in to the community setting, will support immediate completion of the CRAFT after assessment and allow the outcome to be electronically communicated to professionals and teams directly and in real time. With the use of these types of devices becoming more common within the NHS and other health organisations, this may be a vision realised in the near future.

9.4 Further Research

This research in itself will enable further research in the field of acute mental health crisis to take place. Prior to the development of the CRAFT there were no existing valid and reliable assessment measures that could be used to complete research in the crisis population, for example to trial therapeutic interventions or approaches. The development of the CRAFT provides researchers interested in the field of crisis with an appropriate measurement tool to complete research and therefore extend the evidence base for crisis, which will ultimately support CRHTs to deliver more effective high quality interventions.

In terms of further research directly related to the CRAFT, the first step for further research would be to complete a much larger scale pilot with the 66 item global overall CRAFT measure to obtain at least 1000 datasets for analysis. This data set would be specific to the 66 item CRAFT measure and could be scored using the refined individual item rating-scales developed through the Rasch analysis. This would act as a confirmatory piece of research.

Another reason for completing a much larger scale pilot would be to obtain data from a much larger demographic by completing a trial of the measure in a number of crisis teams across the country. As explained earlier (Chapter 1, section 1.2.4.2), the Rasch model supports the outcomes of analysis to be more generalizable due to the separation of the item and person indices. However, it is logical to assume that there will still be some influence of the local population even if this is at a much lower level than from the traditional statistical approaches and therefore it would be wise to collect a sample that could be taken to be more representative of the general population.

Differential Item Functioning (DIF) looks at whether or not item responses are different between different samples of individuals, for example between males and females. If this happens it would suggest that the outcomes of the measure will differ depending on the different groups the person belongs to rather than their position along the construct. It would be expected that where a measure is functioning well it should give the same outcome level on the construct for individuals who are at that same level despite differences such as age, gender, etc. This will be an important understanding for future development of the crisis measure developed through this research for the purposes of further refinement of the measure and for more precise and tailored application but at these very early formative stages of measurement development the focus is on obtaining an item list and item rating scale to accurately represent the crisis construct.

With the collection of a larger sample, the data could be analysed again using the Rasch techniques outlined in this research but with a confirmatory approach to confirm the final item pool. Once the item pool had been confirmed techniques embedded in Structural Equation Modelling could be applied to uncover the underlying structure of the CRAFT measure with the aim of further developing the scoring system to take into account the features of the model, for example any mediating factors or the influence of a bi-factor model on the overall score obtained.

The validity of the measure was assured through the approach adopted in this research. However, it was not possible to assess the predictive validity of the CRAFT before the measure and the cut-offs had been developed. Therefore, one of the aims of future research should be to assess the predictive validity of the CRAFT. The inter-rater reliability of the measure was shown to be weak and the sample too small to provide any meaningful outcomes. It would be expected that a sample of at least 22 to meet the criteria for a large effect size (Cohen, 1992) should be collected using a standardised approach to training on the tool in order to obtain a more accurate and representative inter rater reliability statistic. It would be hoped that with a more standardised and comprehensive approach to training and with an adequate sample the inter-rater reliability outcomes would improve.

The use of the flexible rating scale applied in this research was not directly investigated. It was agreed by the focus groups that the option of using a range score would be a clinically meaningful and useful approach to scoring items due to the changeable nature of individuals experiencing crisis. It would be interesting to research the use of this range score further to try and understand what it represents when used clinically by professionals. For example, it may be hypothesised that the use of the range score will decrease as a patient recovers from crisis and therefore their presentation starts to stabilise. If this hypothesis was shown to be accurate, it may be that the range score represents some meaningful information that would be useful for painting the clinical picture of the patient. For example, it may show that where the range score is used more regularly, the patient is less predictable, which indicates a more acute presentation that would be difficult to manage in the community. In contrast, an individual where only single point scores are used suggests a more stable presentation, which is more predictable in nature. Understanding the use of the range score may provide a powerful clinical tool for further understanding an individual's crisis presentation.

Finally, Rasch analysis is a powerful clinical tool in that it has the capacity to identify the items that are least likely to be scored and therefore highlights items that can act as 'warning signals' to the user. For example, if the item "predictability" was shown to be a 'warning signal' i.e. the least likely item to be scored as causing concern by Rasch analysis, this would indicate that any individual who scores on this item is likely to score highly on any other item in the measure as a *cause for concern*. This ultimately makes the scoring of other items redundant in terms of treatment decision as it could be assumed that this person will score highly on all other items. Rasch has the ability to identify these items and this information can be used as part of Computer Adaptive Testing which uses information provided by Rasch analysis to ensure that only the items necessary to obtain an accurate impression of outcome are administered based on probability. This approach to the administration of the measure would significantly reduce completion time for the measure by effectively reducing the number of items scored, which in turn would further support speedy treatment decision making. This would offer significant value and efficiency to crisis team members by significantly reducing the time taken for acute cases to be identified for intensive support.

Overall, the development of this measure required a return to basics to build a strong foundation on which to build the CRAFT. This was achieved with the support of the patients themselves, mental health professionals working in the field and academic guidance from the research team and peers. It is believed that the CRAFT will provide a clinically meaningful means to measurement that holds particular strengths for clinical application and for the purposes of research by providing an appropriate, population specific measurement tool. With the establishment of this measure, it is hoped that further research can now be completed in this field to develop and improve crisis intervention and treatment. The aim here is to put mental health professionals in the position to support individuals in acute mental health crisis to move away from an experience of danger and towards an experience of opportunity (as encapsulated in the Chinese symbol for crisis). Crisis has the potential to be a growth experience, but without a comprehensive understanding of what crisis is and without the means to measurement, the development of appropriate and helpful treatment and intervention has been stunted. It is hoped that this research has opened the door to further research by providing an accurate and clinically meaningful approach to measuring a presentation as complex, changeable and distressing as acute mental health crisis.

References

Aguilera, D. C. (1998). *Crisis Intervention, Theory and Methodology*. St Louis: Mosby.

Ajzen, I. & Fishbein, M. (1980). *Understanding Attitudes and Predicting Social Behaviour.* NJ: Prentice Hall.

Allen, M. J. & Yen, W. M. (1979). *Introduction to measurement theory.* Monterey, CA.: Brooks/Cole.

Altman, D. G. (1991). *Practical Statistics for Medical Research.* London: Chapman & Hall.

Altman, D. G. & Bland, J. M. (1994). Statistics Notes: Diagnostic tests 3: receiver operating characteristic plots. *British Medical Journal*, 309, 188. Accessed at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2540706/pdf/bmj00449-0054.pdf

Anastasi, A. (1976). *Psychological testing*. (4th ed.) New York: MacMillan.

Andrew, L.B., & Ovsiew, F. (2009). The Psychiatric Interview and The Mental Status Examination. In Bourgeois, J.A., Hales, R.E., Young, J.S., & Yudofsky, S.C. *The American Psychiatric Publishing Board Guide for Psychiatry*. American Psychiatric Publishing. Arlington. Accessed at: http://books.google.co.uk/books?id=RFazteXMaj8C&pg=PR16&lpg=PR16&dq=The+American+Psychiatric+Publishing+Board+Guide+for+Psychiatry&source=bl&ots=OOLADz9MQd&sig=_IOmYgjNUkafm5ZB9B8HpDZC5pA&hl=en&sa=X&ei=DX4uT7vnA6ah0QWN8umsCA&ved=0CF8Q6AEwBw#v=onepage&q=The%20American%20Psychiatric%20Publishing%20Board%20Guide%20for%20Psychiatry&f=false

Andrich, D. A. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.

Antony, M. M. (2004). *Handbook of Assessment and Treatment Planning for Psychological Disorders*. New York: Guilford Press.

Alsuwaidan, M.T., Kucyl, A., Law, C.W.Y., & McIntyre, R.S. (2009). Exercise and Bipolar Disorder: A Review of Neurobiological Mediators. *Neuromolecular Medicine*, *11* (4) 328-336

Atkinson, M. J. & Lennox, R. D. (2006). Extending basic principles of measurement models to the design and validation of patient reported outcomes. *Health and Quality of Life Outcomes, 4*, 1-12. DOI:10.1186/1477-7525-4-65

Akiskal, H.S. (2008). The Mental Status Examination. *The Medical Basis of Psychiatry, 1*. 3-16. DOI: 10.1007/978-1-59745-252-6_1

Baca-Garcia, E., Diaz-Sastre, C., Resa, E. G., Blasco, H., Conesa, D. B., Saiz-Ruiz, J., & Leon, J. (2004). Variables associated with hospitalization decisions by emergency psychiatrists after patients suicide attempt. *Psychiatric Services, 55*, 792-797. DOI: 10.1176/appi.ps.55.7.792

Bambauer, K. Z., Locke, S. E., Aupont, O., Mullan, M. G., & McLaughlin, T. J. (2005). Using the Hospital Anxiety and Depression Scale to screen for depression in cardiac patients. *General Hospital Psychiatry*, *27*, 275-284. DOI: http://dx.doi.org/10.1016/j.genhosppsych.2005.03.002

Baron, J. (2000). Thinking and Deciding. (3rd ed.) USA: Cambridge University Press.

Bartlett, M. S. (1951). A further note on tests of significance in factor analysis. *British Journal of Psychology, 4*, 1-2. DOI: 10.1111/j.2044-8317.1951.tb00299.x

Bartley, W.W. (III). (1987). Philosophy of biology versus philosophy of physics. In Radnitzky, G., Bartley, W.W. (III), editors. *Evolutionary Epistemology, Rationality, and the Sociology of Knowledge*. Open Court; La Salle, IL. P.7-46.[accessed through Googlebooks – books.google.co.uk]

Beck, A. T., Ward, C., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An Inventory for Measuring Depression. *Archives of General Psychiatry, 4*, 561-571.

Beck, A. T., Kovacs, M., & Weissman, A. (1979). Assessment of suicidal ideation: The scale for suicide ideation. *Journal of Consulting and Clinical Psychology, 47*, 343-352. DOI: 10.1037/0022-006X.47.2.343

Beck, A. T. (1986). Hopelessness as a predictor of eventual suicide. In J.J.Mann & M. Stanley (Eds.), *Psychology and Suicidal Behaviour: Annals of the New York Academy of Sciences* (pp. 90-96).

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory* II (BDI-II). San Antonio: Psychology Corporation.

Bengelsdorf, H., Levy, L. E., Emerson, R. L., & Barile, F. A. (1984). A crisis triage rating scale. Brief dispositional assessment of patients at risk for hospitalisation. *The Journal of nervous and mental disease., 172,* 424-430.DOI: 10.1097/00005053-198407000-00009

Berka, K. (1983). *Measurement. Its Concepts, Theories and Problems.* Holland: Reidel.

Berwick, V., Cheek, L., & Ball, J. (2004). Statistics review 13: Receiver operating characteristic curves. *Critical Care, 8*, 508-512.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In M.D.Dunnette & L. M. Hough (Eds.), *Multilevel theory, research, and methods in organizations.* (pp. 349-381). San Francisco: Jossey-Bass.

Bond, T. G. &. F. C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. (2nd ed.) London: Lawrence Erlbaum Associates.

Bonneson M.E.& Hartsough D.M. (1987). Development of the Crisis Call Outcome Rating Scale. *Journal of Consultant Clinical Psychology, 55,* 627-633.

Bonynge, E. R. & Thurber, S. (2008). Development of Clinical Ratings for Crisis Assessment. *Brief Treatment and Crisis Intervention, 8,* 304-312. DOI: 10.1093/brief-treatment/mhn017

Bornstein, R.F. (2011) Toward a process-focused model of test score validity: improving psychological assessment in science and practice. *Psychological Assessment, 23* (2): 532-544. DOI: 10.1037/a0022402.

Brandt, S. (2008). Evolution of a Rasch model including subdimensions. In *Issues and Methodologies in Large-Scale Assessments. IERI Monograph Series*. Volume 1. International Association for the Evaluation of Educational Achievement (IEA) and Educational Testing Service (ETS).

Brimblecombe, N. (2008). Assessment and management of risk. In S.Johnson, J. Needle, & J. P. &. T. G. Bindman (Eds.), *Crisis Resolution and Home Treatment in Mental Health* (pp. 111-120). New York: Cambridge University Press.

Burns, L.R., Bradley, E.H., & Weiner, B.J. (2012). *Shortell and Kaluzny's Health Care Management: Organization Design and Behaviour,* Sixth Edition. Cengage Learning, Delmar.

Campbell, D. T. (1990). The Meehlian Corroboration-Verisimilitude theory of science. *Psychological Inquiry, 1,* 142-147. DOI: 10.1207/s15327965pli0102_2

Campbell, D.T. & Fiske, D.W. (1959). Convergent and discriminant validation by the multitraitmultimethod matrix. Psychological Bulletin, 56 (2), 81-105. DOI: 10.1037/h0046016

Cannon, W. B. (1932). *The Wisdom of the Body* . New York: Norton.

Cano, S. J., Barrett, L. E., Zajicek, J. P., & Hobart, J. C. (2011). Beyond the reach of traditional analyses: using Rasch to evaluate the DASH in people with multiple sclerosis. *Multiple Sclerosis, 17*, 214-222. DOI: 10.1177/1352458510385269

Caplan, G. (1964). *Principles of preventive psychiatry*. New York: Basic Books.

Caplan, G. (1961). *An approach to community mental health.* New York: Basic Books.

Carmines, E. & & Zeller, R. (1979). *Reliability and Validity Assessment, Quantitative Applications in the Social Sciences.* (Vols. 17) Beverly Hills: Sage Publications.

Cattell, J. M. (1886). The time it takes to see and name objects. Mind, 11, 63-65. Accessed at: http://www.jstor.org/stable/2247157.

Cattell, R. B. (1966). The scree test for number of factors. *Multivariate Behavioural Research*, *1*, 245-276. DOI: 10.1207/s15327906mbr0102_10

Cattell, R. B. & Vogelmann, S. (1977). A comprehensive trial of the scree and KG criteria for determining the number of factors. *Journal of Educational Measurement, 14,* 289-325. DOI: 10.1207/s15327906mbr1203_2

Cerel, J. & Campbell, F.R. (2010) Suicide Survivors Seeking Mental Health Services: A Preliminary Examination of the Role of an Active Postvention Model. Suicide and Life Threatening *Behavior. 38* (1), 30-34. DOI: 10.1521/suli.2008.38.1.30

Charmaz, K. (2003). Grounded Theory. In J.Smith (Ed.), *Qualitative Psychology: A practical guide to research methods.* (pp. 81-110). London: Sage.

Chioqueta, A. P. & Stiles, T. C. (2007). The relationship between psychological buffers, hopelessness, and suicidal ideation: Identification of protective factors. *The Journal of Crisis Intervention and Suicide Prevention., 28*, 67-73. DOI: 10.1027/0227-5919.28.2.67.

Chiovitti, R. F. & Piran, R. N. N. (2003). Rigour and grounded theory research. *Journal of Advanced Nursing., 44,* 427-435. DOI: 10.1046/j.0309-2402.2003.02822.x

Christensen, K. J., Multhaup, K. S., Nordstrom, S., & Voss, K. (1991). A cognitive battery for dementia: Development and measurement characteristics. *Journal of Consulting and Clinical Psychology, 3,* 168-174.

Cianci, R. & Gambrel, P. A. (2003). Maslow's hierarchy of needs: Does it apply in a collectivist culture. *Journal of Applied Management and Entrepreneurship, 8,* 143-161. Accessed at: http://www3.tjcu.edu.cn/wangshangketang/lyxgl/yuedu/21.pdf

Clark, L. A. & Watson, D. (1995). Constructing validity: basic issues in objective scale development. *Psychological Assessment, 7,* 309-319.

Cohen, J. (1992). A Power Primer. *Psychological Bulletin, 112,* 155-159. DOI: 10.1037/0033-2909.112.1.155.

Cohen, J. A. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37-46. Accessed at: http://web.njit.edu/~mendonca/statcomp/papers/Cohen_nominal_scales_1960.pdf

Colman, A. M. (2006). *Oxford Dictionary of Psychology*. (2nd ed.) Oxford Paperback Reference.

Colton, D. A., Gao, X., Harris, D. J., Kolen, M. J., Martinovich-Barhite, D., & Wang.T. (1997). Reliability issues with performance assessments: A collection of papers. Iowas City, IA:ACT

Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consultant Clinical Psychology, 56,* 754-761. DOI: 10.1037/0022-006X.56.5.754.

Comrey, A. L. & & Lee, H. B. (1992). *A first course in factor analysis.* (2nd ed.) New Jersey: Erlbaum. Accessed at: http://books.google.co.uk/books?hl=en&lr=&id=Xv1OupfjBDIC&oi=fnd&pg=PR9&dq=Comrey,+A.+L.+%26+%26+Lee,+H.+B.+(1992).+A+first+course+in+factor+analysis.+&ots=g5SyLz10VY&sig=FLrMNOioBVP7ZxAYz06HgfX_cyc

Cook, J. D., Hepworth, S. J., Wall, T. D., & Warr, P. B. (1981). *The experience of work.* San Diego: Academic Press.

Cook, L. L., Eignor, D. R., & Taft, H. L. (1988). A comparative study of the effects of recency of instruction on the stability of IRT and conventional item parameter estimates. *Journal of Educational Measurement, 25*, 31-45. DOI: 10.1111/j.1745-3984.1988.tb00289.x

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334.DOI: 10.1007/BF02310555

Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302. DOI: 10.1037/h0040957

DeMars, C. (2010). *Item Response Theory*. New York: Oxford University Press.

DeMars, C. (2006). Application of the bi-factor multidimensional item response theory model to testlet-based tests. *Journal of Educational Measurement, 43,* 145-168.

Department of Health. *The NHS Plan: A Plan for Investment, a Plan for Reform*. 2000.

Department of Health (2001). *The Mental Health Policy Implementation Guide: Crisis Resolution/Home Treatment Teams* London.

Department of Health. *Equity and excellence: Liberating the NHS.* 2010.

DeVellis, R. F. (2006). Classical Test Theory. Medical Care, 44, 50-59. ISSN: 0025-7079/06/4400-0050

Domino, G. & Domino, M. L. (2006). *Psychological Testing. An Introduction*. New York: Cambridge University Press.

Drogin, E.Y., Dattilio, F.M., Sadoff, R.L. & Gutheil, T.G. (2011). *Handbook of Forensic Assessment. Psychological and Psychiatric Perspectives*. John Wiley & Sons. New Jersey.

Dufresne, T. (2011). *Beyond the Pleasure Principle. Sigmond Freud*. Broadview Press. London. Accessed at: http://books.google.co.uk/books?hl=en&lr=&id=cZsbC1Z-09QC&oi=fnd&pg=PA7&dq=psychic+energy+freud+&ots=Qtjz0bAs7g&sig=yun5fdMbXOX5kAbV2kZPR6VQmgE#v=onepage&q=psychic%20energy%20%28Freud&f=false

Dunn, L. M. (2001). Mental Health Act Assessments; does a community treatment team make a difference? *International Journal of Social Psychiatry, 47,* 1-19. DOI: 10.1177/002076400104700201

Durkheim, E. (1952). *Suicide. A Study in Sociology.* New York: The Free Press.

Eaton, Y. & Ertl, B. (2000). The comprehensive crisis intervention model of Community intervention model of Community Integration, Inc. Crisis Services. In *Crisis intervention handbook: Assessment, treatment and research* . (2 ed., pp. 453-486). New York: Oxford University Press.

Erikson, E. H. (1963). *Childhood and society* . New York: Norton.

Erikson, E. H. (1968). *Identity: Youth and Crisis.* New York: Norton.

Fawcett, J., Scheftner, W. A., Fogg, L., Clark, D. C., Young, M. A., Hedeker, D., et al. (1990). Time-related predictors of suicide in major affective disorder. *American Journal of Psychiatry, 147*(9), 1189-1194.

Field, A. (2011). *Discovering Statistics Using SPSS.* London: Sage.

Fisher, W. (1992). The Rasch debate: validity and revolution in education measurement. In M.Wilson (Ed.), *Objective measurement: theory into practice.* New Jersey: Ablex.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76,* 378-382. DOI: 10.1037/h0031619.

Fletcher, B. (C). (2003). FIT: A new framework for stress & health. In M.J.e.a.Schabrach (Ed.), *Handbook of Work & Health Psychology* (pp. 549-568). John Wiley.

Fletcher, B. (C)., Hanson, J., Pine, K. J., & Page, N. (2011). FIT - Do Something Different: A new psychological intervention tool for facilitating weight loss. *Swiss Journal of Psychology, 70,* 25-34. DOI: 10.1024/1421-0185/a000035

Fletcher, B. (C), Kupshik, G. A., Uprichard, S., Shah, S., & Nash, A. (2008) Eating disorders and concurrent psychopathology: a reconceptualisation of clinical need through Rasch Analysis. *European Journal of Eating Disorders,16,* 191–198. DOI: 10.1002/erv.833

Ford, K. & & McGlynn, P. (2006). Risk Management. In The Sainsbury Centre for Mental Health (Ed.), *Crisis Resolution and Home Treatment.* A Practical Guide London.

Forsyth, R., Saisangjan, U., & Gilmer, J. (1981). Some empirical results related to the robustness of the Rasch model. *Applied Psychological Measurement, 5* (2), 175-186. DOI: 10.1177/014662168100500203

Gangestad, S. W. & Thornhill, R. (1998). The analysis of fluctuating asymmetry redux: The robustness of parametric statistics. *Animal Behaviour, 55*, 497-501. DOI: http://dx.doi.org/10.1006/anbe.1997.0610

Gask, L., Dorwick, C., Klinkman, M., & Gureje, O. (2009). Diagnosis and classification of mental illness: a view from primary care. In Gask, L., Lester, H., Kendrick, T., & Peveler, R. *Primary Care Mental Health*. RCPsych Publications. London.

George, D. & Mallery, P. (2003). *Discovering Statistics using SPSS.* (3rd ed.) Boston: Allyn & Bacon.

Ghiselli, E. E., Campbell, J. P., & Zedek, S. (1981). *Measurement Theory for the Behavioural Sciences*. San Francisco: W.H.Freeman.

Gibbons, R.D., Bock, R.D., Hedeker, D., Weiss, D.J., Segawa, E., Bhaumik, D.K., Kupfer, D.J., Frank, E., Grochocinski, V.J., & Stover, A. (2007). Full-information item bi-factor analysis of graded response data. *Applied Psychological Measurement, 31*, 4-19. DOI: 10.1177/0146621606289485

Gilliland, B. E. &. J. R. K. (1993). *Crisis intervention strategies* . (2 ed.) Pacific Grove: C.A.: Brooks/Cole.

Glass, C. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions of analysis of variance and covariance. *Annual Mathematic Statistics, 29,* 885-889. DOI: 10.3102/00346543042003237

Glover, G., Arts, G., & Babu, K. S. (2006). Crisis resolution/home treatment teams and psychiatric admission rates in England. *British Journal of Psychiatry, 189*, 441-445. DOI: 10.1192/bjp.bp.105.020362

Golan, N. (1978). *Treatment in Crisis Situations*. New York: Free Press.

Golan, N. (1969). When is a client in crisis? *Social Casework., 50,* 389-394.

Goldberg, D. & Williams, P. (1991). *A User's Guide to the General Health Questionnaire.* GHQ. Windsor, UK: NFER-Nelson.

Goldberg, D. P. & Hillier, V. F. (1979). A scaled version of the General Health Questionnaire. *Psychological Medicine, 9*, 139-145. DOI: 10.1017/S0033291700021644

Gorsuch, R. L. (1983). *Factor Analysis.* (2nd ed.) NJ: Lawrence Erlbaum Associates.

Grob, G. N. (1992). Mental health policy in America: myths and realities. *Health Affairs, 11*, 7-22. DOI: 10.1377/hlthaff.11.3.7

Grover, K.E., Green, K.L., Petit, J.W. Monteith, L.L., Garza, M.J., & Venta, A. (2009). Problem solving moderates the effects of life event stress and chronic stress on suicidal behaviors in adolescence. *Journal of Clinical Psychology, 65*, 1281- 1290. DOI: 10.1002/jclp.20632. DOI: 10.1007/BF02289162

Guilford, J. P. (1954). *Psychometric Methods*. (2nd ed.) New York: McGraw & Hill.

Guilford, J. P. (1965). *Fundamental Statistics in Psychology and Education.* (4th ed.) New York: McGraw & Hill.

Guttman, L. (1954). Some necessary conditions for common factor analysis. *Psychometrika, 19,* 149-161.

Hambleton, R. K. & Rovinelli, R. J. (1986). Assessing the dimensionality of a set of test items. *Applied Psychological Measurement, 10*, 287-302. DOI: 10.1177/014662168601000307

Harrison, D. A. (1986). Robustness of parameter estimation to violations to the unidimensionality assumption. *Journal of Educational Statistics, 11,* 91–115. DOI: 10.3102/10769986011002091

Hatcher, L. & Stepanski, E. (1994). *A Step-by-Step Approach to Using the SAS System for Univariate and Multivariate Statistics*. Cary, NC: SAS Institute Inc.

Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and Commitment Therapy: Model, processes and outcomes. *Behaviour Research and Therapy, 44*, 1-25. DOI: http://dx.doi.org/10.1016/j.brat.2005.06.006

Haynes, S. N., Richard, D. R., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*, 238-247. DOI : 10.1037/1040-3590.7.3.238

Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item Response Theory and Health Outcomes Measurement in the 21st Century. Medical Care, 38 (9), 1128-1142.

Henson, R. K. & Roberts, J. K. (2006). Use of exploratory factor analysis in published research. *Educational and Psychological Measurement, 6,,* 393-416.

Hirsch, J.K., & Conner, K.R. (2006). Dispositional and explanatory style optimism as potential moderators of the relationship between hopelessness ad suicidal ideation. *Suicide & Life-Threatening Behavior, 36,* 661-667. DOI: 10.1521/suli.2006.36.6.661

Hobart, J. C., Cano, S. J., Zajicek, J. P., & Thompson, A. J. (2007). Rating scales as outcome measures for clinical trials in neurology: problems, solutions and recommendations. *Lancet Neurology, 6,* 1094-1105. DOI: http://dx.doi.org/10.1016/S1474-4422(07)70290-9

Hobbs, M. (1984). Crisis intervention in theory and practice: a selective review. *British Journal of Medical Psychology, 57*, 23-24. DOI: 10.1111/j.2044-8341.1984.tb01577.x

Hoff, L. A. (2001). *People in Crisis. Clinical Health and Public Health Perspectives*. San Francisco: Jossey-Bass. A Wiley Company.

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

Hoult, J. & Cotton, M. A. (2008). Assessment of Crisis. In S.Johnson, J. Needle, J. P. Bindman, & G. Thornicroft (Eds.), *Crisis Resolution and Home Treatment in Mental Health* (pp. 97-110). Cambridge: Cambridge University Press.

Humphreys, L. G. & Ilgen, D. R. (1969). Note on a criterion for the number of common factors. *Educational and Psychological Measurement*, *29,* 578. DOI: 10.1177/001316446902900303

Ip, E.H. (2010). Empirically indistinguishable multidimensional IRT and locally dependent unidimensional item response models. *British Journal of Mathematical and Statistical Psychology, 63,* 395-416. DOI: 10.1348/000711009X466835

Jacobson, N. & Greenley, D. (2001). What is recovery? A conceptual model and explication. *Psychiatric Services, 52,* 482-485. DOI: 10.1176/appi.ps.52.4.482

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

James, K. J. & Gilliland, B. E. (2001). *Crisis Intervention Strategies.* Pacific Grove: Brook/Cole.

Janis, I. L. & Mann, L. (1997). *Decision making: A psychological analysis of conflict, choice, and commitment*. New York: Free Press.

Janosik, E. H. (1994). *Crisis counselling: A contemporary approach.* Boston: Jones and Bartlett.

Johnson, S., Nolan, F., Pilling, S., Sandor, A., Hoult, J., McKenzie, N., White, I. R., Thompson, M., & Bebbington, P. (2005). Randomised controlled trial of acute mental health care by a crisis resolution team: the north Islington crisis study. *British Medical Journal, 331,* 599. DOI: 10.1136/bmj.38519.678148.8F

Johnson, S. & Thornicroft, G. (2008). The development of crisis resolution and home treatment teams. In S.Johnson, J. Needle, J. P. Bindman, & G. Thornicroft (Eds.). *Crisis Resolution and Home Treatment in Mental Health.* Cambridge: Cambridge University Press.

Johnson, S., Needle, J., Bindman, J. P., & Thornicroft, G. (2008). *Crisis Resolution and Home Treatment in Mental Health.* Cambridge: Cambridge University Press.

Johnson, J.; Wood, A.M.; Gooding, P.; Taylor, P.J. & Tarrier, N. (2011). Resilience to Suicidality: The buffering hypothesis. *Clinical Psychology Review, 31.* 563-591. DOI: http://dx.doi.org/10.1016/j.cpr.2010.12.007

Joiner, T. E., Brown, J. S., & Wingate, L. R. (2005). The psychology and neurobiology of suicidal behaviour. *Annual Review of Psychology, 56,* 287-314. DOI: 10.1146/annurev.psych.56.091103.070320

Jones, W. L. (1968). The A-B-C method of crisis management. *Mental Hygiene, 57,* 87-89. Supplied by the British Library.

Kabat-Zinn, J. (2003). Mindfulness-based interventions in context: Past, Present, and Future. *Clinical Psychology: Science and Practice, 10,* 144-156. DOI: 10.1093/clipsy.bpg016

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20,* 141-151. DOI: 10.1177/001316446002000116

Kaiser, H. F. (1961). A note on Guttman's lower bound for the number of common factors. *Multivariate Behavioural Research, 1,* 249-276.

Kane, M.T. (2001). Current concerns in validity theory. *Journal of Educational Measurement, 38,* 319-342. DOI: 10.1111/j.1745-3984.2001.tb01130.x

Kanel, K. (1998). The ABC model of crisis intervention. *Psychotherapy Integration, 1,* 1-8.

Kass, R. A. & Tinsley, H. E. A. (1979). Factor Analysis. *Journal of Leisure Research, 11,* 120-138.

Kelly, G. A. (1955). *The Psychology of Personal Constructs.* New York: Norton.

Kendall, P. C. & Grove, W. M. (1988). Normative comparisons in therapy outcome. *Behavioral Assessment, 10,* 147-158. Accessed at: http://www.psych.umn.edu/faculty/grove/043normativecomparisonsinpsychotherapy.pdf

Kenny, D. A. (1979). *Correlation and causality.* New York: Wiley.

Keown, P., Tacchi, M. J., Niemiec, S., & Hughes, J. (2007). Changes to mental healthcare for working age adults; impact of a crisis team and an assertive outreach team. *Psychiatric Bulletin, 31,* 288-292. DOI: 10.1192/pb.bp.106.012054

Keyes, C. L. M. (2000). Subjective Change and its Consequences for Emotional Well-Being. *Motivation and Emotion, 24,* 67-84. DOI: 10.1023/A:1005659114155

Keyes, C. L. M. & Ryff, C. D. (2000). Subjective change and mental health: A self-concept theory. *Social Psychology Quarterly, 63,* 264-279. URL: http://www.jstor.org/stable/2695873

Keyes, C. L. M. (2002). The mental health continuum: from languishing to flourishing in life. *Journal of Health and Social Research, 43,* 207-222. Available at: http://www.jstor.org/stable/3090197.

Keyes, C. L. M. (2005). Mental illness and/or mental health? Investigating axioms of the complete state model of health. *Journal of Consulting and Clinical Psychology, 73,* 539-548. DOI: 10.1037/0022-006X.73.3.539

Keyes, C. L. M. (2007). Promoting and protecting mental health as flourishing: A complementary strategy for improving national mental health. *American Psychologist, 62,* 95-108. DOI: 10.1037/0003-066X.62.2.95

Kiseliova, D. & Kiseliova, A. (2004). *Diagnostics of Mathematical Skills. Research Monograph*. Shiauliai University Publishing.

Klein, D. C. & Lindemann, E. (1961). Preventative Intervention in Individual and Family Crisis Situations. In *Prevention of Mental Disorders in Children.* New York: Basic Books.

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

Landy, F. J. (1986). Stamp collecting versus science: validation as hypothesis testing. *American Psychologist, 41,* 1183-1192. DOI: 10.1037/0003-066X.41.11.1183

Lazarus, R. S. & Folkman, S. (1984). *Stress, Appraisal and Coping.* New York: Springer.

Lazarus, R. S. (1993). Coping theory and research: past, present, and future. *Psychosomatic Medicine*, *55*, 234-247. Accessed at: http://www.emotionalcompetency.com/papers/coping%20research.pdf

Lazarus, R. S. &.Folkman, F. S. (1984). *Stress, appraisal, and coping*. New York: Springer.

Lebreton, J. M., Burgess, J. R. D., Kaiser, R. B., Atchley, E. K., & James, L. R. (2003). The restriction of variance hypothesis and inter rater reliability and agreement: Are ratings from multiple sources really dissimilar? *Organizational Research Methods*, *6*, 80-128. DOI: 10.1177/1094428102239427

Lebreton, J. M. & Senter, J. L. (2008). Answers to 20 questions about inter rater reliability and interrater agreement. *Organizational Research Methods, 11,* 815-852. DOI: 10.1177/1094428106296642.

Lee, Y. (2004). Examining passage-related local item dependence (LID) and measurement construct using Q3 statistics in an EFL reading comprehension test. *Language Testing, 21,* 74-100. DOI: 10.1191/0265532204lt260oa

Leventhal, H., Brissette, I., & Leventhal, E. A. (2003). The common-sense model of self-regulation of health and illness. In L.D.Cameron & H. Leventhal (Eds.), *The self regulation of health and illness behaviour*. London: Routledge.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 22,* 55.

Linacre, J. M. (2002). Optimsing Item Rating Scale Category Effectiveness. *Journal of Applied Measurement*, *3,* 85-106.

Linacre, J. M. (2006). Demarcating Category Intervals. *Rasch Measurement Transcripts*, *19*, 1041-1043. ISSN: 1051-0796.

Lindemann E (1956). The meaning of crisis in individual and family. *Teachers Collection Records*, *57*, 310.

Lindemann, E. (1944). Symptomatology and management of acute grief. *American Journal of Psychiatry, 101*, 141-148.

Linehan, M. M. (1987). Dialectical Behaviour Therapy for borderline personality disorder: Theory and method. *Bulletin of the Menninger Clinic, 51*, 261-276.

Linehan, M.M., Comtois, K.A., & Ward-Ciesielski. (2011). Assessing and managing risk with suicidal individuals. *Cognitive and Behaviour Practice*. Article in press. Available at: http://www.sciencedirect.com/science/article/pii/S1077722911000599 DOI: 10.1016/j.cbpra.2010.11.008.

Linn, R. L. (1980). Issues of validity for criterion-referenced measures. *Applied Psychological Measurement, 4*, 561. DOI: 10.1177/014662168000400407

Lissitz, R. W. & Green, S. B. (1975). Effect of the number of scale points on reliability. A Monte Carlo approach. *Journal of Applied Psychology, 60*, 10-13. DOI: 10.1037/h0076268

Loevinger, J. (1957). Objective tests as instruments of psychological theory. Psychological *Rep.Monograph Supplement, 3*, 635-694.

Loewenthal, K. M. (2001). *An Introduction to Psychological Tests and Scales*. (2nd ed.) Padstow: Psychology Press.

Lopez de Lara, C., Dumais, A., Rouleau, G., Lesage, A., Dumont, M., Chawky, N., Alda, M., Benkelfat, C., & Turecki, G. (2006). STin2 variant and family history of suicide as significant predictors of suicide completion in major depression. *Biological Psychiatry, 59,* 114-120. DOI: http://dx.doi.org/10.1016/j.biopsych.2005.06.021

Lord, F. M. & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores.* Reading: Addison-Wesley.

Lowry, R. (2011). *Clinical Calculator* [Computer software]. Accessed at http://faculty.vassar.edu/lowry/clin1.html.

MacCorquodale, K. & Meehl, P.E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review, 55* (2), 95-107. DOI: 10.1037/h0056029

MacLeod, A.K., Rose, G.S., & Williams, J.M. (1993). Components of hopelessness about the future in parasuicide. *Cognitive Therapy and Research, 17,* 441-455. DOI: 10.1007/BF01173056

Maslow, A. H. (1943). A Theory of Human Motivation. *Psychological Review, 50,* 370-396. DOI: 10.1037/h0054346

Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47,* 149-174. DOI: 10.1007/BF02296272

Masters, G. N. & Wright, B. D. (1997). The Partial Credit Model. In W.J.Van der Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory*. (pp. 101). New York: Springer.

Masters, G. N. & Wright, B. D. (1984). The essential process in a family of measurement models. *Psychometrika*, *49,* 529-544. DOI: 10.1007/BF02302590

McGlynn, P. (2006). *Crisis Resolution and Home Treatment: A practical guide.* U.K.: Nuffield Press, The Sainsbury Centre for Mental Health.

McIntire, S. A. & Miller, L. A. (2007). *Foundations of Psychological Testing*. A Practical Approach. (2nd ed.) London: Sage Publications.

McIver, J. P. & Carmines, E. G. (1981). *Unidimensional Scaling*. London: Sage Publications.

Meehl, P. E. (1945). The dynamics of "structured" personality tests. *Journal of Clinical Psychology, 1,* 296-303. DOI: 10.1002/1097-4679(194510)1:4<296::AID-JCLP2270010410>3.0.CO;2-#

Messer, S. B. & Wampold, B. E. (2002). Let's Face Facts: Common factors are more potent than specific therapy ingredients. *Clinical Psychology: Science and Practice, 9*, 21-25. DOI: 10.1093/clipsy.9.1.21

Messick, S. (1989). Validity. In R.L.Linn (Ed.), *Educational Measurement* (pp. 13-103). New York: Macmillan.

Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741-749. DOI: 10.1037/0003-066X.50.9.741

Miller, L. A., McIntire, S. A., & Lovler, R. L. (2010). *Foundations of Psychological Testing. A Practical Approach*. London: Sage Publications.

Millon, T., Millon, C., Davis, R. & Grossman, S. (1997). *Millon Clinical Multiaxial Inventory-III*. Pearson.

Mirrlees-Black, C. (1999). *Domestic Violence: findings from a new British Crime Survey self-completion questionnaire*. Home Office Research Study 191. London.

Mok, M. C. C. & Flynn, M. (2002). Determinants of students' quality of school life: A path model. *Learning Environment Research, 5,* 275-300. DOI: 10.1023/A:1021924322950

Montanelli, R. G. & Humphreys, L. G. (1976). Latent roots of random data correlation matrices with squared multiple correlations on the diagonal: A Monte Carlo study. *Psychometrika, 41,* 341-348. DOI: 10.1007/BF02293559

Morrison, R., & O’Connor, R.C. (2008). A systematic review of the relationship between rumination and suicidality. *Suicide & Life Threatening Behavior, 38,* 523-538. DOI: 10.1521/suli.2008.38.5.523

Morton, A. H. (1937). *And sin no more: social policy and unwed mothers in Cleveland*. Ohio: Ohio State University Press.

Myer RA & Conte C. (2006). Assessment for crisis intervention. *Journal of Clinical Psychology, 62,* 959-970. DOI: 10.1002/jclp.20282

Myer, R. A., Williams, R. C., Ottens, A. J., & Schmidt, A. E. (1992). Crisis assessments: A three-dimensional model for triage. *Journal of Mental Health Counselling, 14*, 137-148.

Myers, K. & Winter, M. C. (2002). Ten-year review of rating scales. Overview of scale functioning, psychometric properties and selection. *Journal of the American Academy of Child and Adolescent Psychiatry. 41*, 114-122.

Mykletun, A. M., Stordal, E., & Dahl, A. A. (2001). Hospital Anxiety and Depression (HAD) scale: factor structure, item analysis and internal consistency in a large population. *The British Journal of Psychiatry, 179*, 540-544. DOI: 10.1192/bjp.179.6.540

National Audit Office. *Helping People Through Mental Health Crisis: The Role of Crisis Resolution and Home Treatment Services*. 2007. London.

Neimeyer, R. A. (2004). Constructions of death and loss: Evolution of a research program. *Personal Construct Theory & Practice, 1,* 8-20. Accessed at: http://www.pcp-net.org/journal/pctp04/neimeyer04.pdf

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

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

Oakley, C. & Malik, A. (2010). *Rapid Psychiatry*. Wiley-Blackwell. Oxford.

O'Connor, R.C., & Noyce, R. (2008). Personality and cognitive processes; Self-criticism and different types of rumination as predictors of suicidal ideation. *Behaviour Research and Therapy. 46*, 392-401. DOI: http://dx.doi.org/10.1016/j.brat.2008.01.007

Olsson, I., Mykletun, A. & Dahl, A.A. (2005). The hospital anxiety and depression rating scale: A cross-sectional study of psychometrics and case finding abilities in general practice. BMC *Psychiatry, 5* (46). 1-7.

Online Corrected Version (2011). *IUPAC, Compendium of Chemical Terminology*. (2nd ed.).

Ouellette, J. A. & Wood, W. (1998). Habit and intention in everyday life: The multiple processes by which past behaviour predicts future behaviour. *Psychological Bulletin, 124,* 54-74. DOI: 10.1037/0033-2909.124.1.54

PASW (formally SPSS) (2010). (Version 18) [Computer software]. IBM.

Pallant, J. F. & Tennant, A. (2007). An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology, 46,* 1-18. DOI: 10.1348/014466506X96931

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Perlmutter R.A.& Jones J.E. (1985). Assessment of families in psychiatric emergencies. *American Journal of Orthopsychiatry, 55,* 130-139.

Plutchik, R. (1980). *Emotions: A Psychoevolutionary Synthesis*. New York: Harper & Row.

Pollock, L.R., & Williams, J.M.G. (2001). Effective problem solving in suicide attempters depends on specific autobiographical recall. *Suicide & Life Threatening Behavior, 31,* 386-396.

Pollock, L.R., & Williams, J.M.G. (2004). Problem solving in suicide attempters. *Psychological Medicine, 34,* 163-167. DOI: 10.1521/suli.31.4.386.22041

Porta, L. F., Franceschini, M., Caselli, S., Cavallini, P., Susassi, S., & Tennant, A. (2011). Unified Balance Scale: an activity-based, bed to community, and aetiology-independent measure of balance calibrated. *Journal of Rehabilitation Medicine*, Epub ahead of print.

Power, A. K. (2009). A Public Health Model of Mental Health for the 21st Century. *Psychiatric Services, 60,* 580-584.

Preacher, K. J. & MacCallum, R. C. (2003). Repairing Tom Swift's Electric Factor Analysis Machine. *Understanding Statistics, 2,* 13-43.

Priester, M.J., & CLum, G.A. (1993b). The problem-solving diathesis in depression, hopelessness, and suicide ideation: A longitudinal analysis. *Journal of Psychopathology and Behavioural Assessment, 15,* 239-254. DOI: 10.1007/BF01371381

Puri, B. K. & Treasden, I. (2010). *Psychiatry. An Evidence-based text.* London: Hodder Arnold.

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement Equivalence: A Comparison of Methods Based on Confirmatory Factor Analysis and Item Response Theory. *Journal of Applied Psychology, 87,* (3): 517-529. DOI: 10.1037//0021-9010.87.3.517.

Rapoport, L. (1962). The state of crisis: some theoretical considerations. *Social Service Review, 36,* 211. URL: http://www.jstor.org/stable/30016886

Rasch, G. (1960). *Probabilistic Models for some Intelligence and Attainment Tests*. Chicago: University of Chicago Press.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4,* 207-230. DOI: 10.3102/10769986004003207

Reckase, M. D. (2009). *Multidimensional Item Response Theory.* London: Springer.

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality Life Resources, 16,* 19-31. DOI: 10.1007/s11136-007-9183-7

Reise, S. P. & Waller N.G. (2009). Item Response Theory in Clinical Measurement. *Annual Review of Clinical Psychology, 5,* 27-48. DOI: 10.1146/annurev.clinpsy.032408.153553

Repper, J. & Perkins, R. (2006). *Social Inclusion and Recovery: A Model for Mental Health Practice*. UK: Biliere Tindall.

Roberts, A. R. (2000). *An overview of crisis theory and crisis intervention. In Crisis intervention handbook: Assessment, treatment, and research* (2 ed.) Oxford University Press. Oxford.

Roberts, A. R. (2002). Assessment, crisis intervention and trauma treatment: The integrative ACT intervention model. *Brief Treatment and Crisis Intervention, 2,* 1-21. Accessed at: http://btci.edina.clockss.org/cgi/content/abstract/2/1/1

Roberts, A. R. (2005). *Crisis Intervention Handbook: Assessment, Treatment, and Research.* Oxford: Oxford University Press.

Roberts, A. R. &. L. S. (2001). Crisis Assessment Tools: The Good, the Bad and the Available. *Brief Treatment and Crisis Intervention, 1,* 17-28. Accessed at: http://btci.stanford.clockss.org/cgi/content/abstract/1/1/17

Rogers, J. (1994). *Validity analysis of the suicide risk assessment checklist.* Akron: University of Akron.

RUMM2030 (2010). [Computer software]. www.rummlab.com.

Ryrie, I., Hellard, L., Kearns, C., Robinson, D., Pathmanathan, I., & Sullivan, D. (1997). Zoning: A system for managing case work and targeting resources in community mental health teams. *Journal of Mental Health, 6,* 515-523. DOI: 10.1080/09638239718608

Schmitt, N. (2011). Uses and abuses of coefficient alpha. *Psychological Assessment, 8,* 350-353. DOI: 10.1037/1040-3590.8.4.350

Seigert, R. J., Jackson, D. M., Tennant, A., & Turner-Stokes, L. (2010). Factor analysis and Rasch analysis of the Zarit Burden Interview for Acquired Brain Injury carer research. *Journal of Rehabilitation Medicine, 42,* 302-309. DOI: http://dx.doi.org/10.2340/16501977-0511

Sharma, S. (2011). *FIT Science for improving family functioning and parental stress*. University of Hertfordshire, UK.

Sifneos P.E (1960). A concept of emotional crisis. *Mental Hygiene, 44,* 169.

Simms, L. J., Casillas, A., Clark, A. C., Watson, D., & Doebbeling, B. N. (2005). Psychometric Evaluation of the Restructured Clinical Scales of the MMPI-2. *Psychological Assessment, 17,* 345-358. DOI: 10.1037/1040-3590.17.3.345

Simms, L. J. & Watson, D. (2007). The construct validation approach to personality scale construction. In R.W.Robins, C. R. Fraley, & R. F. Krueger (Eds.), *Research Methods on Personality Psychology*. New York: The Guildford Press.

Simon, H. (1953). Causal ordering and identifiability. In W.C.Hood & T. C. Koopmans (Eds.), *Studies in Econometric Method*. (pp. 49-74). New York: Wiley.

Slaideu K.A (1990). *Crisis Intervention. A Handbook for Practice and Research*. MA, US: Allyn & Bacon.

Slinde, J. A. & Linn, R. L. (1979a). A note on vertical equating via the Rasch model for groups of quite different ability and tests of quite different difficulty. *Journal of Educational Measurement, 16,* 159-165. DOI: 10.1111/j.1745-3984.1979.tb00097.x

Slinde, J.A. & Linn, R.L. (1979b). The Rasch Model, objective measurement, equating, and robustness. *Applied Psychological Measurement, 3* (4). 437-452. DOI: 10.1177/014662167900300402

Smith, A. B., Wright, E. P., Rush, R., Stark, D. P., Velikova, G., & Selby, P. J. (2006). Rasch Analysis of the Dimensional Structure of the Hospital Anxiety and Depression Scale. *Psycho-Oncology, 15,* 817-827. DOI: 10.1002/pon.1015

Smith, E. V. (2002). Understanding Rasch Measurement: Detecting and Evaluating the Impact of Multidimensionality using Item Fit Statistics and Principal Component Analysis of Residuals. *Journal of Applied Measurement, 3,* 205-231.

Smith, G.T., Fischer, S., Fister, S.M. (2003). Incremental validity principles in test construction. *Psychological Assessment, 15* (4): 467-477. DOI: 10.1037/1040-3590.15.4.467

Smith, R. M. (2000). Fit analysis in latent trait measurement models. *Journal of Applied Measurement, 1,* 199-218.

Spearman, C. (1906). "Footrule" for measuring correlation. *British Journal of Psychology, 2,* 89-108. DOI: 10.1111/j.2044-8295.1906.tb00174.x

Spector, P. E. (1992). A consideration of the validity and meaning of self-report measures of job conditions. In *International review of industrial and organizational psychology* (pp. 123-151). Chichester: Wiley.

Stanimirova, M., Daszykowski, B., & Walczak (2006). Dealing with missing values and outliers in principal component analysis. *Talanta*, *72*, 172-178. DOI: http://dx.doi.org/10.1016/j.talanta.2006.10.011

Stern, P. (1985). Using grounded theory method in nursing research. In N.N.Leininger (Ed.), *Qualitative research methods in nursing*. (pp. 149-160). Orlando, FL.: Grune & Stratton.

Stevens, J. P. (2002). Applied Multivariate Statistics for the Social Sciences. Mahwah, New Jersey: Routledge.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103,* 677-680.

Stone, M. H. & Wright, B. D. (1994). Maximising rating scale information. *Rasch Measurement Transcripts*, *8*, 386. Accessed at: http://www.rasch.org/rmt/rmt83r.htm

Stout, W. (1987). A non-parametric approach for assessing latent trait unidimensionality. *Psychometrika*, *52,* 589-617.

Strauss, A. & Corbin, J. (1990). *Basics of Qualitative Research: Grounded theory procedures and techniques*. London: Sage.

Strauss, M. E. & Smith, G. T. (2009). Construct Validity: Advances in Theory and Methodology. *Annual Review in Clinical Psychology*, *5*, 1-25. DOI: 10.1146/annurev.clinpsy.032408.153639

Strohle, A. (2009). Physical activity, exercise, depression and anxiety disorders. *Journal of Neural Transmission, 116,* (6) 777-784. DOI: 10.1007/s00702-008-0092-x

Suen, H. K. (1990). *Principles of Test Theories*. London: Lawrence Erlbaum Associates.

Sylvia, L.G., Ametrano, R.M., & Nierenberg. (2010) Exercise Treatment for Bipolar Disorder: Potential Mechanisms of Action Mediated through Increased Neurogenesis and Decreased Load. *Psychotherapy and Psychosomatics. 79,* 87-96. DOI: 10.1159/000270916

Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics.* (5th ed.) Boston: Allyn & Bacon.

Tacchi, M.J., & Scott, J. (2008). Strategies for promoting engagement and treatment adherence. In Johnson, S., Needle, J., Bindman, J. P., & Thornicroft, G. *Crisis Resolution and Home Treatment in Mental Health.* Chapter 13. Cambridge: Cambridge University Press.

Taplin, J. R. (1971). Crisis theory: Critique and reformulation . *Community Mental Health Journal, 7,* 13-23. DOI: 10.1007/BF01434728

Tennant, A. & Pallant, J. F. (2006). Unidimensionality matters. *Rasch Measurement Transcripts, 20,* 1048-1051.

Tennant, A., McKenna, S. P., & Hagell, P. H. (2004). Application of Rasch Analysis in the Development and Application of Quality of Life Instruments. *Value in Health, 7,* 22-26.

Tennant, A. & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? *Arthritis & Rheumatism, 57,* 1358-1362. DOI: 10.1002/art.23108.

The Concise Corsini Encyclopedia of Psychology and Behavioural Science. 3rd. 2004. NJ, John Wiley and Sons. Encyclopaedias. Craighead, W. E. and Nemeroff, C. B.

The Department of Health, T. S. O. *The NHS Plan.* 2000.

Thoits, P. A. (1995). Stress, coping, and social support processes: Where are we? What next? *Journal of Health and Social Behaviour, 35,* 53-79. URL: http://www.jstor.org/stable/2626957

Thomas, W.I. (1909). *Source Book for Social Origins*. Chicago: University of Chicago Press.

Thurstone, L. L. (1931). *The reliability and validity of tests.* MI: Edwards Brothers.

Thurstone, L. L. (1936). A new conception of intelligence. *Educational Record, 17,* 441-450.

Tobitt, S. & Kamboj, S. (2011). Crisis resolution/home treatment team workers' understandings of the concept of crisis. *Social Psychiatry Psychiatric Epidemiology, 46,* 671-683. DOI: 10.1007/s00127-010-0234-y

Triandis, H. C. (1977). *Interpersonal Behaviour.* Monterey, C.A.: Brooks/Cole.

Triandis, H. C. (1979). Values, attitudes, and interpersonal behaviour. In H.E.Howe & M. M. Page (Eds.), *Nebraska Symposium on Motivation, Beliefs, Attitudes and Values*. (pp. 195-259).

Trochim, W. M. K. (1989). An Introduction to concept mapping for planning and evaluation. Evaluation and Program Planning, 12, 1-16. DOI: Available online 3 July 2002.DOI: http://dx.doi.org/10.1016/0149-7189(89)90016-5

Truant, G., O'Reilly, R., & Donaldson, L. (1991). How psychiatrists weigh risk factors when assessing suicide risk. *Suicide and life-threatening behaviour, 21*, 106-114. DOI: 10.1111/j.1943-278X.1991.tb00459.x

Tucker, L. R. & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1-10. DOI: 10.1007/BF02291170

Turner, N. E. (1998). The effect of common variance and structure pattern on random data eigenvalues: Implications for the accuracy of parallel analysis. *Educational and Psychological Measurement, 58,* 541-568. DOI: 10.1177/0013164498058004001

Turner, R. J. & Avison, W. R. (1992). Innovations in the Measurement of Life Stress: Crisis Theory and the Significance of Event Resolution . *Journal of Health and Social Behavior, 33,* 36-50. Stable URL: http://www.jstor.org/stable/2136856

Tzeng, O. C. S. (1992). On reliability and number of principal components: Joinder with Cliff and Kaiser. *Perceptual and Motor Skills, 75,* 929-930. DOI: 10.2466/pms.1992.75.3.929

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41,* 321-327. DOI: 10.1007/BF02293557

Wainer, H., Yeomans, K. A., & Golder, P. A. (1982). The Guttman-Kaiser criterion as a predictor of the number of common factors. *The Statistician, 31,* 221-229. Stable URL: http://www.jstor.org/stable/2987988

Ware, J. E., Harris, W. J., Gandek, B., Rogers, B. W., & Reese, P. R. (1997). *MAP-R for Windows: Multitrait/multi-item analysis program* - revised user's guide. Boston, MA: Health Assessment Lab.

Watkins, M. A. (2008). *MacParallel Analysis* [Computer software].

Weisstein, E. W. (2011). *Logistic Equation*. http://mathworld.wolfram.com/LogisticEquation [Announcement posted on the World Wide Web].

Wenzel, A. & Beck, A.T. (2008). A cognitive model of suicidal behaviour: Theory and treatment. *Applied and Preventative Psychology, 12,* 189-201. DOI: http://dx.doi.org/10.1016/j.appsy.2008.05.001,

Westerhof, G. J. & Keyes, L. M. (2010). Mental Illness and Mental Health: Two continua model across the lifespan. *Journal of Adult Development, 17,* 110-119. DOI: 10.1007/s10804-009-9082-y.

Wise, E. A. (2004). Methods for analyzing psychotherapy outcomes: A review of clinical significance, reliable change, and recommendations for future directions. *Journal of Personality Assessment, 82,* 50-59. DOI**:** 10.1207/s15327752jpa8201_10

Wright, B. D. (1999). Common sense for measurement. *Rasch Measurement Translation, 13,* 704-705. Accessed at: http://www.rasch.org/rmt/rmt133h.htm

Wright, B. D. (1999). Fundamental measurement for psychology. In S.E.Embretson & S. L. Hershberger (Eds.), *The New Rules for Measurement: What Every Educator and Psychologist Should Know*. (pp. 65-104). New Jersey: Lawrence Erlbaum Associates.

Wright, B. D. & Masters, G. N. (1982). *Rating Scale Analysis*. Chicago: MESA Press.

Wright, S. S. (1921). Correlation and causation. *Journal of Agricultural Research, 20,* 557-585. Accessed at: http://www.ssc.wisc.edu/soc/class/soc952/Wright/Wright_Correlation%20and%20Causation.pdf

Xitao, F. (1998) Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Educational and Psychological Measurement, 58* (3): 357-381. DOI: 10.1177/0013164498058003001

Yang, S. (2007). *A Comparison of Unidimensional and Multidimensional Rasch Models using Parameter Estimates and Fit Indices when assumption of Unidimensionality is Violated.* Approved PhD Dissertation with the Graduate School of The Ohio State University. Accessed at: http://etd.ohiolink.edu/view.cgi/Yang%20Seungho.pdf?osu1195695378

Yang, B., & Clum, G.A. (1994). Life stress, social support and problem-solving skills predictive of depressive symptoms, hopelessness, and suicide ideation in an Asian student population: A test of a model. *Suicide & Life-Threatening Behavior, 24*, 127-139. DOI: 10.1111/j.1943-278X.1994.tb00797.x

Yerkes, R. M. & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology and Psychology, 18*, 459-482.
DOI: 10.1002/cne.920180503

Zhu, W., Updyke, W. F., & Lewandowski, C. (1997). Post-hoc Rasch analysis of optimal categorisation of an ordered response scale. *Journal of Outcome Measurement, 1,* 286-304. Accessed at: http://www.ncbi.nlm.nih.gov/pubmed/9661725

Zigmond, A. S. & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, *67,* 361-370.

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

10.1207/s15327906mbr1702_5

Zwick, W. R. & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*, 432-442. DOI: 10.1037/0033-2909.99.3.432

Appendix

Appendix 1 – Research Ethics Committee Approval Letter

RECEIVED
11 FEB 2008

NHS

National Research Ethics Service

Bedfordshire Research Ethics Committee
Ambulance Training Centre
Via Location Code Q7
QE11 Hospital
Howlands
Welwyn Garden City
Hertfordshire
AL7 4HQ

Telephone: 01707 362585
Facsimile: 01707 394475

07 February 2008

Miss Nicole Y Stokoe
Research Assisstant
Bedfordshire and Luton Mental Health and Social Care Partnership Trust
Disability Resource Centre, Poynters House
Poynters Road
Dunstable LU5 4TP

Dear Miss Stokoe

Full title of study:	**Enhancing Staff Appraisals of Risk for the Crisis Resolution and Home Treatment Service in Bedfordshire and Luton Developing an Assessment Tool to Objectify Risk**
REC reference number:	**08/H0309/5**

Thank you for your letter of 21 January 2008, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information was considered at the meeting of the Sub-Committee of the REC held on 07 February 2008. A list of the members who were present at the meeting is attached.

Confirmation of ethical opinion

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

Ethical review of research sites

The Committee has designated this study as exempt from site-specific assessment (SSA. There is no requirement for [other] Local Research Ethics Committees to be informed or for site-specific assessment to be carried out at each site.

Conditions of approval

The favourable opinion is given provided that you comply with the conditions set out in the attached document. You are advised to study the conditions carefully.

This Research Ethics Committee is an advisory committee to East of England Strategic Health Authority
The National Research Ethics Service (NRES) represents the NRES Directorate within the National Patient Safety Agency and Research Ethics Committees in England

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Investigator CV		
Protocol	2	21 January 2008
Covering Letter		28 December 2007
Questionnaire: Demographic, Service User	1	21 December 2007
Questionnaire: Demographic, Staff	1	21 December 2007
Letter of invitation to participant	1, Service User	21 December 2007
Letter of invitation to participant	1, Staff Focus Group	21 December 2007
Letter of invitation to participant	1, Service User Meeting	21 December 2007
Participant Information Sheet: Service User Interview	1	21 December 2007
Participant Information Sheet: Staff Interview	1	21 December 2007
Participant Information Sheet: Service User Focus Group	1	21 December 2007
Participant Information Sheet: Staff Focus Group	1	21 December 2007
Participant Consent Form: Service User	1	21 December 2007
Participant Consent Form: Staff	1	21 December 2007
Response to Request for Further Information	Email	21 January 2008
List of Support Agencies	1	21 December 2007
Educational Supervisor's CV		

R&D approval

All researchers and research collaborators who will be participating in the research at NHS sites should apply for R&D approval from the relevant care organisation, if they have not yet done so. R&D approval is required, whether or not the study is exempt from SSA. You should advise researchers and local collaborators accordingly.

Guidance on applying for R&D approval is available from http://www.rdforum.nhs.uk/rdform.htm.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Now that you have completed the application process please visit the National Research Ethics Website > After Review

Here you will find links to the following

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

b) Progress Reports. Please refer to the attached Standard conditions of approval by Research Ethics Committees.
c) Safety Reports. Please refer to the attached Standard conditions of approval by Research Ethics Committees.
d) Amendments. Please refer to the attached Standard conditions of approval by Research Ethics Committees.
e) End of Study/Project. Please refer to the attached Standard conditions of approval by Research Ethics Committees.

We would also like to inform you that we consult regularly with stakeholders to improve our service. If you would like to join our Reference Group please email referencegroup@nationalres.org.uk .

08/H0309/5	**Please quote this number on all correspondence**

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

Yours sincerely

PP.

Mr Ron Driver
Chair

Email: jenny.austin@nhs.net

Enclosures: Standard approval conditions, SL-AC2

Copy to: Dr Gary Kupshik,
Research Governance Approvals Group
Disabilities Resource Centre
Poynters House, Poynters Road
Dunstable, Bedfordshire
LU5 4TP

Appendix 2 – Participant Information Sheet (Staff)

Patient Assessment for the Crisis Resolution and Home Treatment Service in Bedfordshire and Luton

Developing an Assessment Tool to Objectify Risk

Participant Information Sheet (Staff)

You are being invited to take part in a research study looking at how we make decisions about levels of risk that service users present with. Before you decide whether you want to participate or not, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Talk to others about the study if you wish.

Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

About the research

The Crisis Resolution and Home Treatment (CRHT) service was developed to provide rapid assessment of service user's experiencing acute and severe mental illness. This assessment provides the basis upon which the service user's care path is decided, whether to offer intensive input and support, to facilitate and support short stay in crisis beds or to assist in managing an admission when no community alternative is available.

The CRHT service uses a RAG/traffic light system to identify which service users are severe (red), moderate (amber) or recovered (green). To date there are no standardised assessment tools that support staff in to assess service users using the RAG approach.

This research aims to do the following:

1. To develop an assessment tool that will support staff of the CRHT in their assessments of patient's needs.
2. To develop an assessment tool that will make it easier to define the criteria for the RAG levels.
3. To develop an assessment tool that will support staff to decide upon a patient's RAG status.
4. To develop an assessment tool that can be used throughout the intervention to show the progress of the patient through recovery.

Why have I been chosen?

You have been chosen because you have been working within the CRHT for more than 6 months and have experience of working with the RAG system. To help develop the assessment tool it will be useful to look at your experience of the RAG system and how you are currently assessing service user's. All members of staff who have been working for the CRHT for more than 6 months and have experience of assessment will be invited to complete an interview.

Do I have to take part?

No. It is up to you to decide whether or not to take part. If you do, you will be given this information sheet to keep and be asked to sign a consent form. You are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect you in anyway.

What will happen to me if I decide to take part?

You will be invited to attend an interview to look at how assessments are carried out with service users at present. It will be a discussion based around three service users with whom you have or are working with. You will be asked to bring with you three patient cases – one from each of the categories e.g. a patient who is Green, a patient who is Amber and a patient who is Red. Patient names do not need to be disclosed.

You will then be asked to talk briefly about each of the cases and what you felt categorised them as the colour they are under on the RAG system. You will then be invited to discuss further the differences between the red, amber and green categories.

Notes will be taken during the interview in addition to a voice recording. The recording is simply to ensure that all important points discussed within the meeting are noted. If you do not wish to have your interview recorded, please let the interviewer know and the interview will be recorded in note format only.

It is important to note that there are no right or wrong answers. It is useful and important to record your opinions, expertise and 'gut instinct' as well as the practices of working outlined by your manager/team. Your expertise and experience will be essential to developing an appropriate and workable tool.

You are not obliged to answer all or any of the questions, even if you have consented to be interviewed.

The interview should take between 1-2 hours. The research project itself will be completed over a 3 year period and within this time you may be approached a number of times.

What you need to bring

All you need to bring is yourself and three cases/patient's notes – one falling into each of the colours on the RAG system.

Confidentiality and storage of information/data

The interview will be recorded and notes taken. The information will be stored in a locked filing cabinet in the Psychology Department. No personnel other than members of the research team will have access to identifiable records. When information is shared outside of the research team or taken off the premises, any identifiable information will be removed.

Disadvantages/ Benefits of taking part

There are no direct disadvantages from participating in this research project. The possible benefit of participating would be contributing to the development of an assessment tool that could be utilised by the service and therefore staff members.

What if there is a problem?

Any complaint about the way you have been dealt with during the study will be addressed. Please contact Prof. Gary Kupshik, Head of Psychology, Department of Psychology, Disability Resource Centre, Poynters House, Poynters Road, Luton. Tel: 01582 709085 email: gary.kupshik@blpt.nhs.uk

If you experience any problems related to the research that you are happy to discuss directly with the researcher, please contact Nicole Stokoe, Research Assistant using the contact details below.

Contact details

Nicole Stokoe
Research Assistant
Disability Resource Centre, Poynters House, Poynters Road, Luton. Tel: 01582 709085 or 07968242578
nicole.stokoe@blpt.nhs.uk

Prof. Gary Kupshik
Head of Psychology
Disability Resource Centre, Poynters House, Poynters Road, Luton. Tel: 01582 709085
gary.kupshik@blpt.nhs.uk

Appendix 3 – Participant Information Sheet Service User

Enhancing Staff Appraisals of Risk for the Crisis Resolution and Home Treatment Service in Bedfordshire and Luton

Developing an Assessment Tool to Objectify Risk

Participant Information Sheet

You are being invited to take part in a research study looking to develop an assessment tool for use by Crisis Resolution and Home Treatment teams. Before you decide whether you want to participate or not, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Talk to others about the study if you wish.

Please ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

About the research

The Crisis Resolution and Home Treatment (CRHT) service was developed to provide rapid assessment of service user's experiencing acute and severe mental illness. This assessment provides the basis upon which the service user's care path is decided, whether to offer intensive input and support, to facilitate and support short stay in crisis beds or to assist in managing an admission when no community alternative is available.

The CRHT service uses a RAG (Red. Amber. Green)/traffic light system to identify which service users are severe (red), moderate (amber) or recovered (green). To date there are no standardised assessment tools that support staff to assess service users using the RAG approach.

This research aims to do the following:

1. To develop an assessment tool that will support staff of the CRHT in their assessments of patient's needs.
2. To develop an assessment tool that will make it easier to define the criteria for the RAG levels.
3. To develop an assessment tool that will support staff to decide upon a patient's RAG status.
4. To develop an assessment tool that can be used throughout the intervention to show the progress of the patient through recovery.

Why have I been chosen?

You have been chosen because you have had experience of receiving input from and working with the Crisis Resolution and Home Treatment team. You will have personal knowledge and experience about what it is to experience a crisis. Your experience will be useful in supporting the development of the assessment tool for the CRHT and making sure we capture some of your insights into crisis and recovery from crisis.

Do I have to take part?

No. It is up to you to decide whether or not to take part. If you do, you will be given this information sheet to keep and be asked to sign a consent form. You are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect you in anyway.

The research project itself will be completed over a 3 year period and within this time you may be approached a number of times. If you do not wish to be approached again, please contact Nicole Stokoe (details at the bottom of the page) and she will remove your name from the mailing list.

What will happen to me if I decide to take part?

You will be invited to attend an interview to look at how you identified in yourself that you were experiencing crisis and what the signs were to indicate that you were recovering. You will not be obliged to answer any or all of the questions even after you have consented to be interviewed, only the ones you are comfortable to talk about.

Notes will be taken during the interview in addition to a voice recording. The recording is simply to ensure that all the important points discussed within the meeting are noted. If you do not wish to have your interview recorded, please let the interviewer know and the interview will be recorded in note format only.

It is important to note that there are no right or wrong answers. It is useful and important to record your opinions, expertise and 'gut instinct'. Your expertise and experience will be essential to developing an appropriate and workable tool.

The interview should take between 1-2 hours. You may be asked to attend a second interview if it is felt that more time is required. The research project itself will be completed over a 3 year period. You may be approached at other times during this 3 year period and asked to participate in other phases of the research for example in the focus group. If you do not wish to be approached again, please contact Nicole Stokoe (details at the bottom of the page) and she will remove your name from the mailing list.

What you need to bring

You don't need to bring anything with you, the interviewer will be responsible for organising the session.

Confidentiality and storage of information/data

Only members of the research team will be aware of your participation in the research. The CRHT team will not be made aware of your participation unless you yourself decide to speak to them about it. You are free to talk to anyone you wish to about your involvement with the research. Your participation in the research will not affect your statutory rights or access to services in any way.

The interview will be recorded and notes taken. The information will be stored in a locked filing cabinet in the Psychology Department which is the base for the researchers. No personnel other than members of the research team will have access to identifiable records. When information is shared outside of the research team or taken off the premises, any identifiable information will be removed.

Disadvantages/ Benefits of taking part

Talking about your experience of crisis may evoke some uncomfortable memories or feelings. You will be provided with an information sheet at the end of the interview that will provide you with contact details of support agencies that you can get in touch with should these feelings continue. If you feel unhappy at any point during the interview it is important for you to let the researcher know and the interview will be stopped if requested.

What if there is a problem?

Any complaint about the way you have been dealt with during the study will be addressed. Please contact Prof. Gary Kupshik, Head of Psychology, Department of Psychology, Disability Resource Centre, Poynters House, Poynters Road, Luton. Tel: 01582 709085 email: gary.kupshik@blpt.nhs.uk

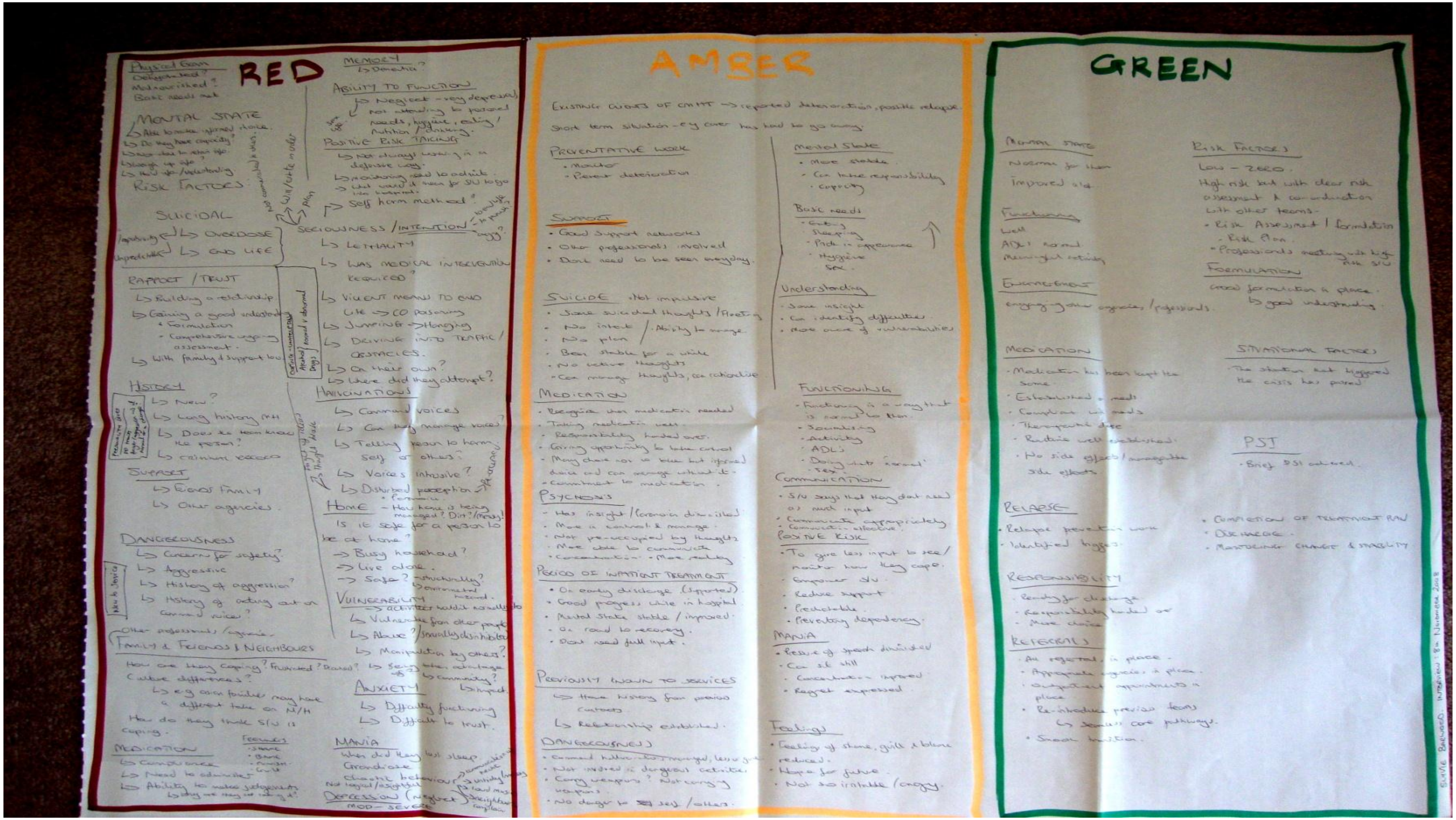
If you experience any problems related to the research that you are happy to discuss directly with the researcher, please contact Nicole Stokoe, Research Assistant using the contact details below.

Contact details
Nicole Stokoe
Research Assistant
Disability Resource Centre, Poynters House, Poynters Road, Luton. Tel: 01582 709085 or 07968242578

nicole.stokoe@blpt.nhs.uk

Prof. Gary Kupshik
Head of Psychology
Disability Resource Centre, Poynters House, Poynters Road, Luton. Tel: 01582 709085
gary.kupshik@blpt.nhs.uk

Appendix 4 - Staff Interview FlipChart Record (exmple)



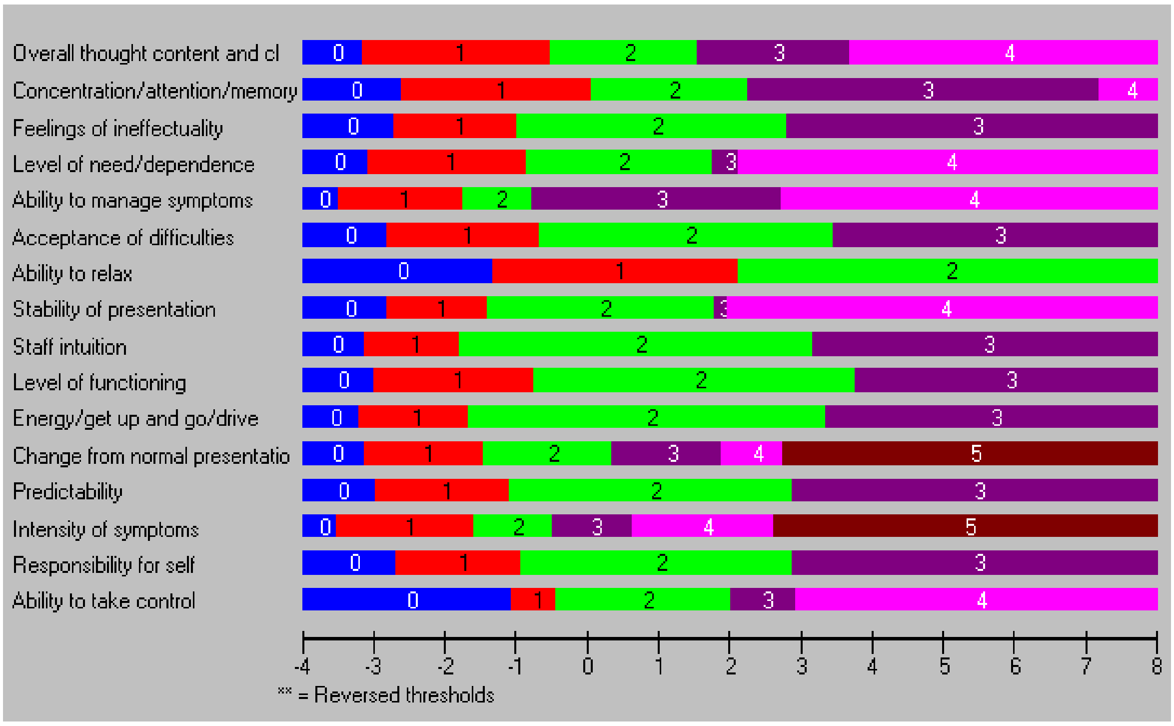
Appendix 5 – Principal Component Analysis using Oblique Rotation

Component Correlation Matrix

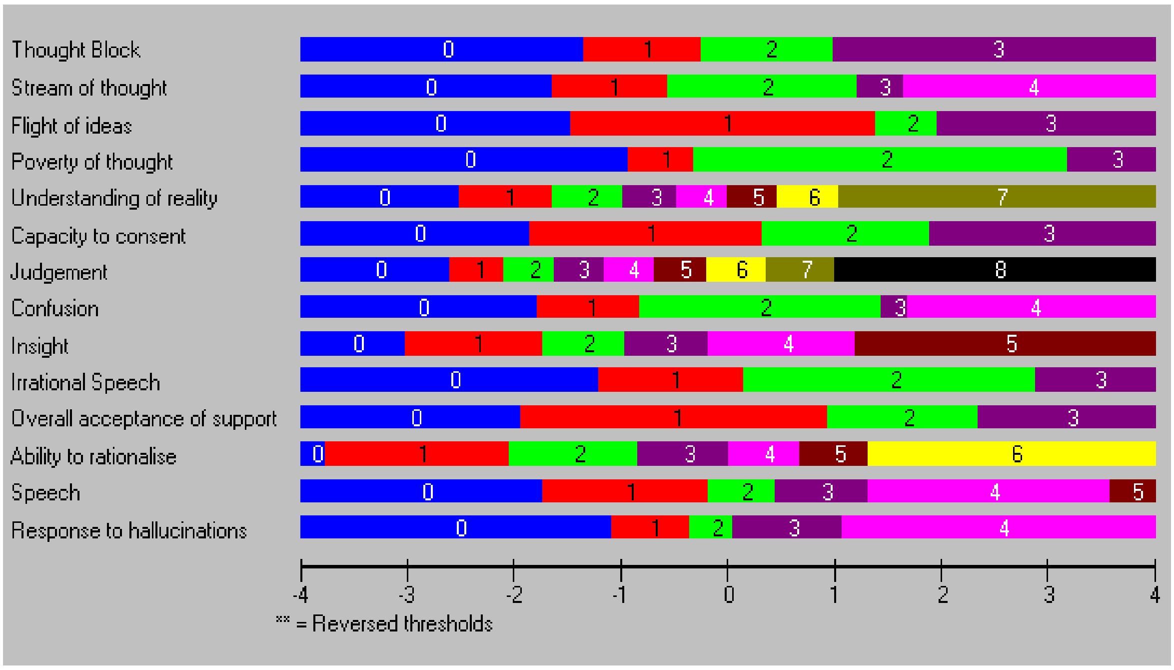
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1.000	.230	.276	.229	-.303	-.265	-.162	-.270	.254	-.091	.195	.220	.252	.147	.106	.212	.214	.265	.244
2	.230	1.000	.048	.123	-.200	-.308	-.306	-.107	.250	-.111	.093	.291	.021	.144	.156	.131	.052	.267	.205
3	.276	.048	1.000	.130	-.297	-.070	-.141	-.226	.154	-.073	.232	.120	.311	.082	.073	.177	.205	.160	.190
4	.229	.123	.130	1.000	-.269	-.211	-.175	-.258	.185	-.150	.164	.209	.320	.158	.135	.139	.168	.156	.237
5	-.303	-.200	-.297	-.269	1.000	.355	.186	.232	-.219	.159	-.159	-.199	-.252	-.162	-.108	-.294	-.275	-.296	-.215
6	-.265	-.308	-.070	-.211	.355	1.000	.270	.102	-.270	.123	-.154	-.220	-.170	-.168	-.098	-.126	-.172	-.348	-.209
7	-.162	-.306	-.141	-.175	.186	.270	1.000	.218	-.233	.184	-.098	-.240	-.123	-.118	-.160	-.097	-.052	-.306	-.223
8	-.270	-.107	-.226	-.258	.232	.102	.218	1.000	-.232	.173	-.161	-.134	-.289	-.149	-.138	-.199	-.209	-.221	-.171
9	.254	.250	.154	.185	-.219	-.270	-.233	-.232	1.000	-.163	.115	.147	.147	.129	.062	.226	.153	.196	.193
10	-.091	-.111	-.073	-.150	.159	.123	.184	.173	-.163	1.000	-.043	-.089	-.108	-.155	-.078	-.130	-.098	-.100	-.117
11	.195	.093	.232	.164	-.159	-.154	-.098	-.161	.115	-.043	1.000	.207	.240	.123	.102	.012	.295	.238	.219
12	.220	.291	.120	.209	-.199	-.220	-.240	-.134	.147	-.089	.207	1.000	.174	.157	.125	.096	.170	.244	.246
13	.252	.021	.311	.320	-.252	-.170	-.123	-.289	.147	-.108	.240	.174	1.000	.121	.106	.119	.224	.207	.219
14	.147	.144	.082	.158	-.162	-.168	-.118	-.149	.129	-.155	.123	.157	.121	1.000	.103	.117	.166	.178	.216
15	.106	.156	.073	.135	-.108	-.098	-.160	-.138	.062	-.078	.102	.125	.106	.103	1.000	.036	.077	.129	.130
16	.212	.131	.177	.139	-.294	-.126	-.097	-.199	.226	-.130	.012	.096	.119	.117	.036	1.000	.172	.099	.170
17	.214	.052	.205	.168	-.275	-.172	-.052	-.209	.153	-.098	.295	.170	.224	.166	.077	.172	1.000	.253	.282
18	.265	.267	.160	.156	-.296	-.348	-.306	-.221	.196	-.100	.238	.244	.207	.178	.129	.099	.253	1.000	.283
19	.244	.205	.190	.237	-.215	-.209	-.223	-.171	.193	-.117	.219	.246	.219	.216	.130	.170	.282	.283	1.000
20	.087	.219	.092	.033	-.048	-.114	-.150	.019	.156	-.103	-.049	.061	-.013	.040	.047	.187	-.097	.005	.066
21	.130	.161	.045	.238	-.237	-.158	-.249	-.204	.128	-.181	.080	.143	.077	.164	.155	.152	.129	.206	.187
22	-.227	-.136	-.300	-.074	.239	.202	.186	.120	-.206	.167	-.189	-.160	-.146	-.152	-.065	-.178	-.195	-.224	-.217

Appendix 6 – Subscale Item Threshold Maps

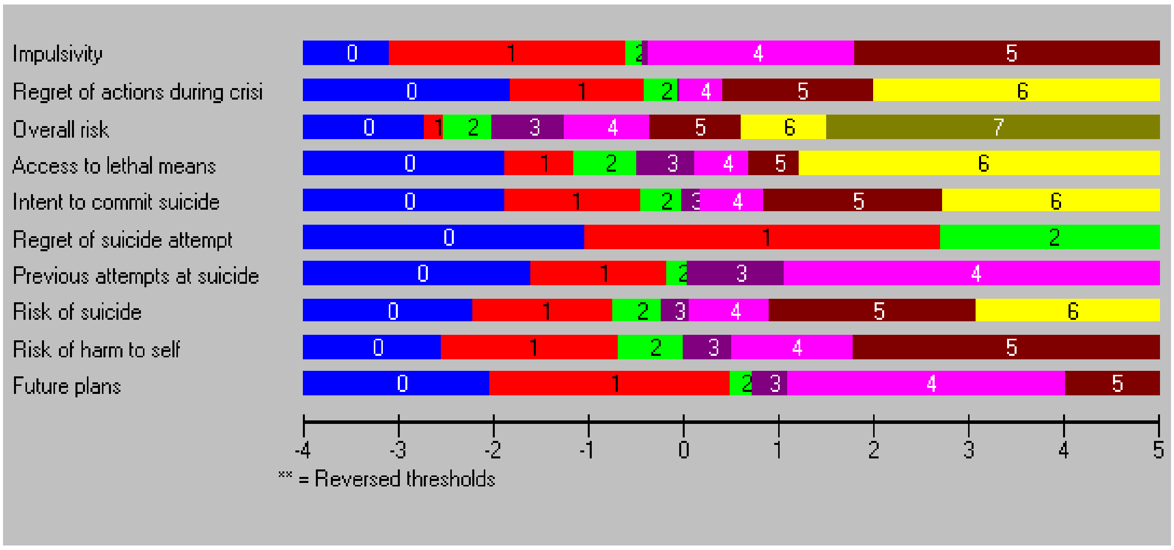
Subscale 1 – Crisis Recovery Indicators



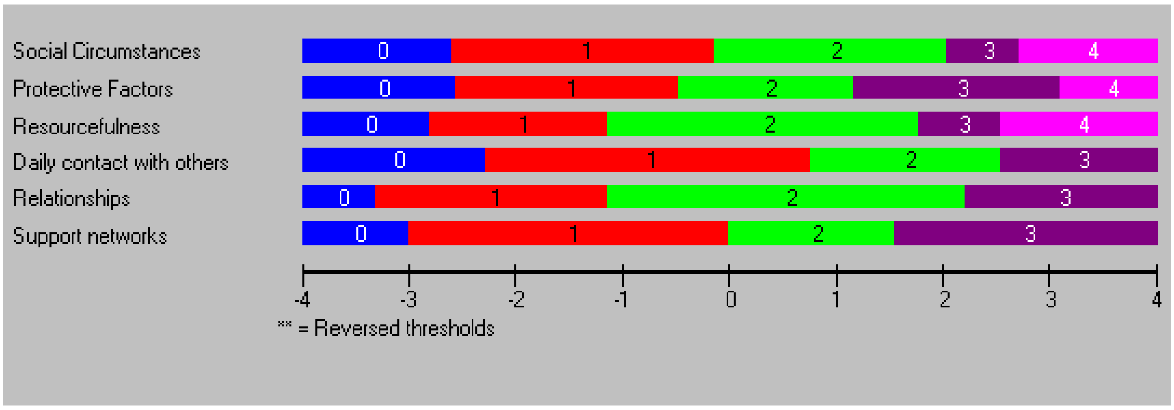
Subscale 2 – Adaptive Decision Making



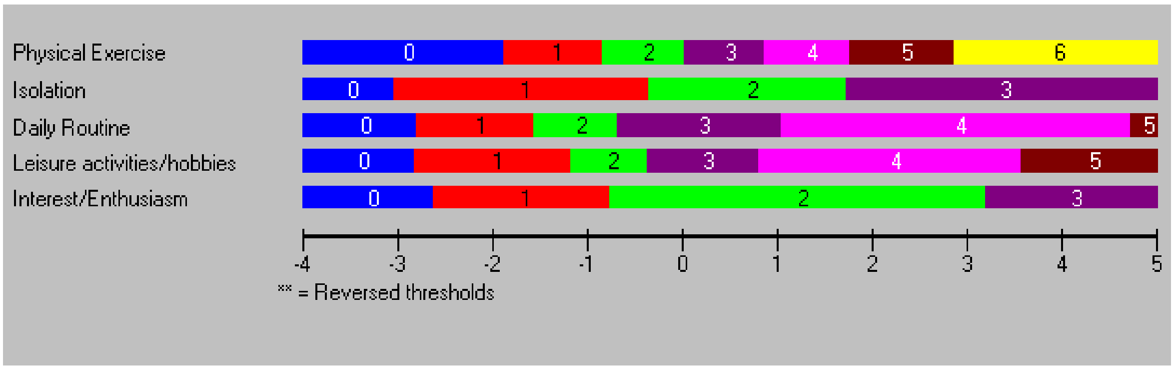
Component 3 – Risk of Harm to Self



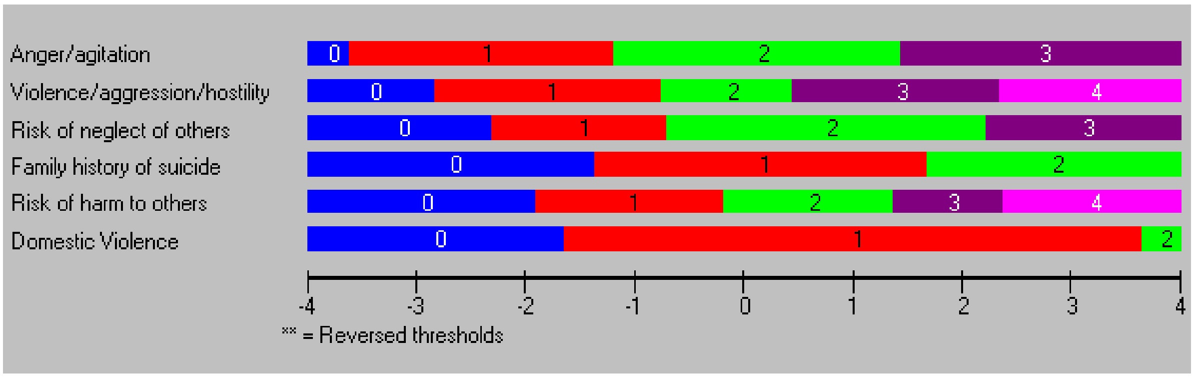
Component 4 – Mediating Factors



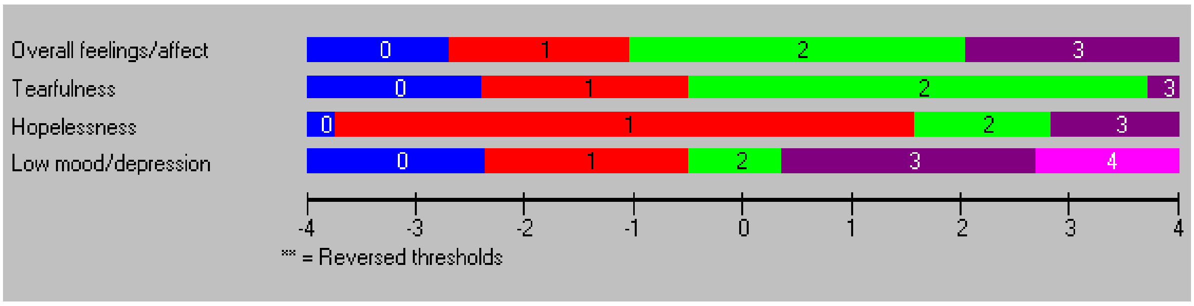
Component 5 – Daily Routine



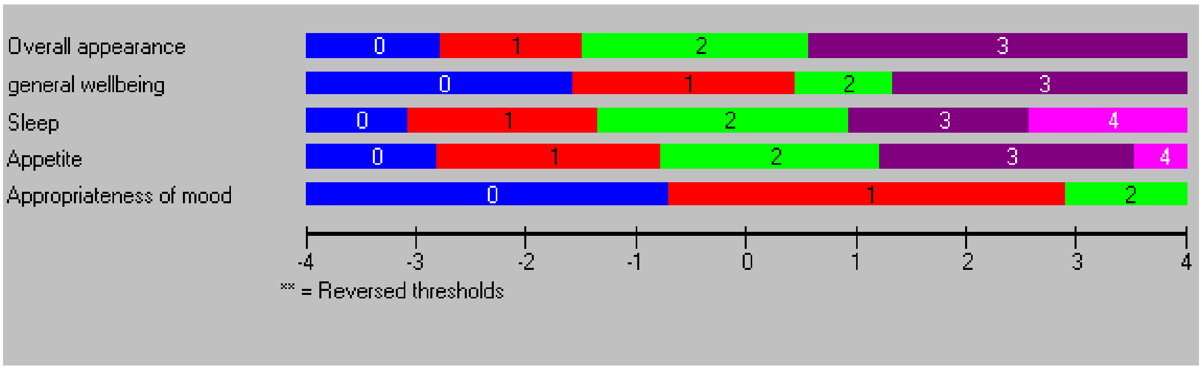
Component 6 – Risk of Harm to Others



Component 7 – Feelings/Affect



Component 8 – Basic Functioning



Appendix 7 - Residual Correlation Matrices for Item Local Dependency

Pearson r>+0.3 highlighted in pink. Pearson r >-0.3 highlighted in green

Subscale 1 – Crisis Recovery Indicators

Item	Overall Thought Content & Clarity	Concentration	Feelings of ineffectuality	Level of need	Ability to manage symptoms	Acceptance of difficulties	Ability to Relax	Stability of presentation	Staff intuition	Level of functioning	Energy/Get up and go	Change from normal presentation	Predictability	Intensity of Symptoms	Responsibility for self	Ability to take control
Overall Thought Content and Clarity	1.000															
Concentration	0.131	1.000														
Feelings of ineffectuality	-0.081	-0.070	1.000													
Level of need	-0.110	-0.107	-0.004	1.000												
Ability to manage symptoms	-0.148	-0.133	-0.062	-0.006	1.000											
Acceptance of difficulties	-0.042	0.028	-0.055	-0.118	-0.159	1.000										
Ability to Relax	-0.064	-0.086	-0.129	-0.117	0.206	0.090	1.000									
Stability of presentation	0.005	-0.082	-0.148	-0.081	-0.175	-0.070	-0.175	1.000								
Staff intuition	-0.072	-0.274	0.044	-0.085	-0.145	-0.138	-0.115	0.120	1.000							
Level of functioning	-0.104	-0.067	0.039	-0.021	-0.021	-0.028	0.010	-0.185	-0.056	1.000						
Energy/Get up and go	-0.192	0.037	-0.096	-0.047	-0.102	-0.008	-0.050	-0.183	-0.100	0.009	1.000					
Change from normal presentation	0.009	-0.069	-0.273	-0.148	-0.067	-0.162	0.005	-0.086	0.006	-0.056	0.067	1.000				
Predictability	0.112	-0.155	-0.130	-0.159	-0.202	0.057	-0.180	0.092	-0.129	-0.080	-0.144	0.059	1.000			
Intensity of symptoms	0.043	0.029	-0.082	-0.266	-0.068	-0.094	0.001	-0.011	0.082	-0.155	-0.115	0.013	-0.036	1.000		
Responsibility for self	-0.189	0.006	-0.133	0.090	-0.093	0.013	-0.056	-0.035	-0.245	-0.056	-0.089	-0.155	0.136	-0.146	1.000	
Ability to take control	-0.309	-0.131	0.085	0.148	-0.004	-0.086	-0.054	-0.091	-0.013	0.058	0.030	-0.159	-0.189	-0.196	0.067	1.000

Subscale 2 – Adaptive Decision Making

Item	Thought Block	Stream of Thought	Flight of Ideas	Poverty of Thought	Understanding of Reality	Capacity to Consent	Judgement	Confusion	Insight	Irrational Speech	Overall Acceptance of Support	Ability to Rationalise	Speech	Response to hallucinations/delusions
Thought Block	1.000													
Stream of Thought	0.320	1.000												
Flight of Ideas	0.151	0.399	1.000											
Poverty of Thought	0.215	0.134	0.209	1.000										
Understanding of Reality	-0.155	-0.108	-0.118	-0.204	1.000									
Capacity to Consent	0.043	-0.072	-0.013	0.071	-0.051	1.000								
Judgement	-0.253	-0.322	-0.289	-0.179	-0.133	0.010	1.000							
Confusion	0.142	-0.020	0.129	0.044	-0.056	-0.044	-0.195	1.000						
Insight	-0.250	-0.208	-0.229	-0.147	-0.207	-0.056	0.045	-0.199	1.000					
Irrational Speech	-0.069	0.070	0.197	-0.101	-0.042	-0.082	-0.273	0.040	-0.273	1.000				
Overall Acceptance of Support	-0.035	-0.057	-0.117	-0.021	-0.157	-0.031	-0.111	-0.106	0.031	0.012	1.000			
Ability to Rationalise	-0.198	-0.054	-0.222	-0.158	-0.211	-0.171	-0.079	-0.199	0.215	-0.116	-0.094	1.000		
Speech	0.019	-0.061	-0.006	0.077	-0.116	0.077	-0.194	-0.040	-0.167	0.069	0.004	-0.098	1.000	
Response to Hallucinations/delusions	-0.094	-0.006	-0.025	-0.039	0.011	-0.102	-0.218	0.002	-0.253	0.350	0.112	-0.223	-0.016	1.000

Subscale 3 – Risk of Harm to Self

Item	Impulsivity	Regret if actions during crisis	Overall risk	Access to lethal means	Intent to commit suicide	Regret of suicide attempt	Previous attempts at suicide	Risk of suicide	Risk of harm to self	Future plans
Impulsivity	1.000									
Regret of actions during crisis	-0.005	1.000								
Overall risk	-0.065	-0.256	1.000							
Access to lethal means	-0.015	-0.229	-0.057	1.000						
Intent to commit suicide	-0.255	-0.266	-0.130	0.010	1.000					
Regret of suicide attempt	-0.043	0.145	-0.339	-0.165	-0.049	1.000				
Previous attempts at suicide	-0.200	-0.124	-0.186	-0.171	-0.172	0.127	1.000			
Risk of suicide	-0.265	-0.271	-0.106	-0.111	0.187	0.006	-0.014	1.000		
Risk of harm to self	-0.116	-0.135	-0.074	-0.236	-0.056	-0.095	-0.157	0.008	1.000	
Future plans	-0.219	0.026	-0.252	-0.120	0.006	0.002	-0.104	-0.094	0.090	1.000

Subscale 4 - Mediating Factors

Item	Social Circumstances	Protective factors	Resourcefulness	Daily contact with others	Relationships	Support Networks
Social Circumstances	1.000					
Protective factors	-0.107	1.000				
Resourcefulness	-0.294	-0.182	1.000			
Daily contact with others	-0.094	-0.223	-0.179	1.000		
Relationships	-0.158	-0.385	-0.294	-0.246	1.000	
Support Networks	-0.333	0.039	-0.281	-0.021	0.362	1.000

Subscale 5 – Daily Routine

Item	Physical Exercise	Isolation	Daily Routine	Leisure activities/hobbies	Interest/Enthusiasm
Physical Exercise	1.000				
Isolation	-0.261	1.000			
Daily Routine	-0.466	-0.158	1.000		
Leisure activities/hobbies	-0.322	-0.257	-0.062	1.000	
Interest/Enthusiasm	-0.166	-0.113	-0.367	-0.193	1.000

Subscale 6 – Risk of Harm to Others

Item	Anger/Agitation	Violence/hostility/ aggression	Risk of neglect of others	Family history of suicide	Risk of harm to others	Domestic Violence
Anger/Agitation	1.000					
Violence/hostility/aggression	0.187	1.000				
Risk of neglect of others	-0.270	-0.186	1.000			
Family history of suicide	-0.311	-0.292	-0.103	1.000		
Risk of harm to others	-0.288	-0.024	-0.131	-0.020	1.000	
Domestic Violence	-0.246	-0.153	-0.020	0.011	0.003	1.000

Subscale 7 – Feelings and Affect

Item	Overall Feelings	Tearfulness	Hopelessness	Low mood/depression
Overall Feelings	1.000			
Tearfulness	-0.208	1.000		
Hopelessness	-0.110	-0.081	1.000	
Low mood/depression	-0.206	-0.143	-0.137	1.000

Subscale 8 – Basic Functioning

Item	Overall appearance	General wellbeing	Sleep	Appetite	Appropriateness of mood
Overall appearance	1.000				
General wellbeing	-0.170	1.000			
Sleep	-0.369	-0.208	1.000		
Appetite	-0.322	-0.187	-0.210	1.000	
Appropriateness of mood	-0.063	-0.237	-0.216	-0.053	1.000

Appendix 8 – Item Fit Residuals for 66 Items of the 8 Subscales

Residuals greater than +/-3.0 highlighted in orange

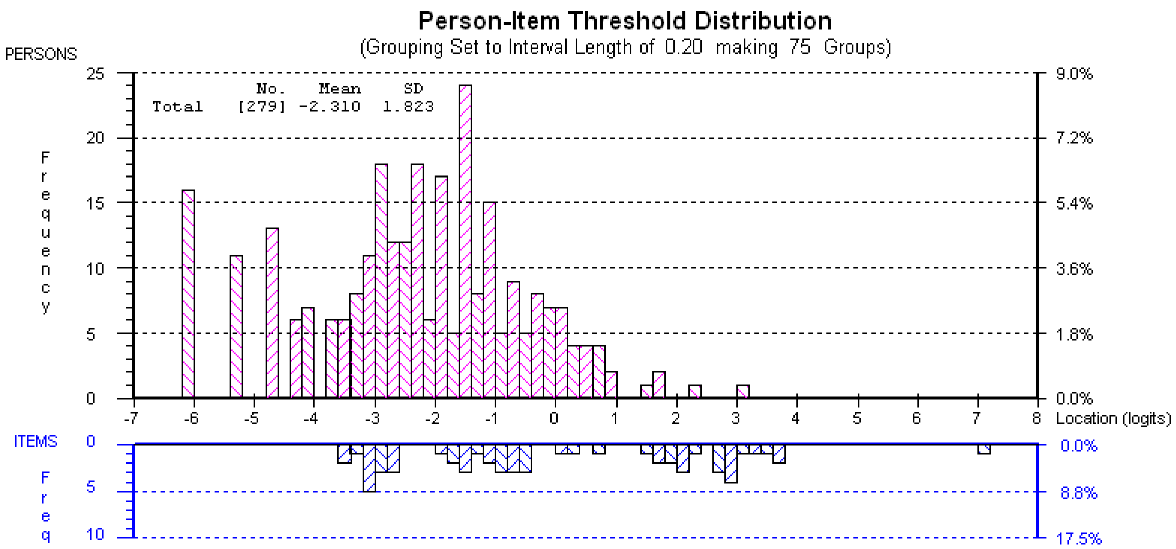
Item residuals greater than +/- 2.5 highlighted in green

					Fit Statistics	
Subscale	**Item no.**	**Item Label**	**Location**	**SE**	**Residual**	**Chi-square**
Component 1 – Recovery Indicators	1	Overall thought content and clarity	0.394	0.105	-0.180	5.250
	2	Concentration	1.722	0.111	0.900	5.071
	3	Feelings of ineffectuality	-0.295	0.105	2.152	10.411
	4	Level of need	-0.016	0.101	0.760	4.604
	5	Ability to manage symptoms	-0.824	0.087	0.631	3.484
	6	Acceptance of difficulties	-0.006	0.109	-0.221	6.546
	7	Ability to relax	0.389	0.139	0.510	4.343
	8	Stability of presentation	-0.119	0.097	-0.372	7.765
	9	Staff Intuition	-0.589	0.104	1.278	4.589
	10	Level of functioning	0.014	0.110	-1.213	7.555
	11	Energy/get up and go	-0.513	0.105	0.265	3.419
	12	Change from normal presentation	0.085	0.089	-1.015	2.086
	13	Predictability	-0.393	0.105	-0.723	4.614
	14	Intensity of symptoms	-0.471	0.082	**-2.538**	8.187
	15	Responsibility for self	-0.240	0.105	0.381	5.837
	16	Ability to take control	0.862	0.105	0.417	3.193
Component 2 – Adaptive Decision Making	17	Thought block	-0.201	0.102	-1.235	7.788
	18	Stream of thought	0.166	0.097	-0.788	8.841
	19	Flight of ideas	0.628	0.127	-1.180	18.00
	20	Poverty of thought	0.651	0.112	-0.676	15.567
	21	Understanding of reality	-0.588	0.063	**3.036**	14.365
	22	Capacity to consent	0.126	0.112	**-3.234**	19.125
	23	Judgement	-0.875	0.055	**2.810**	17.536
	24	Confusion	0.127	0.096	0.466	2.167
	25	Insight	-0.939	0.072	1.092	5.936
	26	Irrational speech	0.617	0.116	-2.492	8.618
	27	Overall acceptance of support	0.448	0.120	0.069	10.879
	28	Ability to rationalise	-0.773	0.072	1.894	7.013
	29	Speech	0.687	0.094	-0.043	5.591
	30	Response to hallucinations/delusions	-0.074	0.091	1.253	5.522
Component 3 – Risk of Harm to Self	31	Impulsivity	-0.536	0.066	1.356	7.981
	32	Regret of actions during crisis	0.015	0.063	2.265	12.219
	33	Overall risk	-0.964	0.053	-0.972	1.305
	34	Access to lethal means	-0.243	0.060	-0.426	4.655
	35	Intent to commit suicide	0.237	0.066	-2.103	8.773

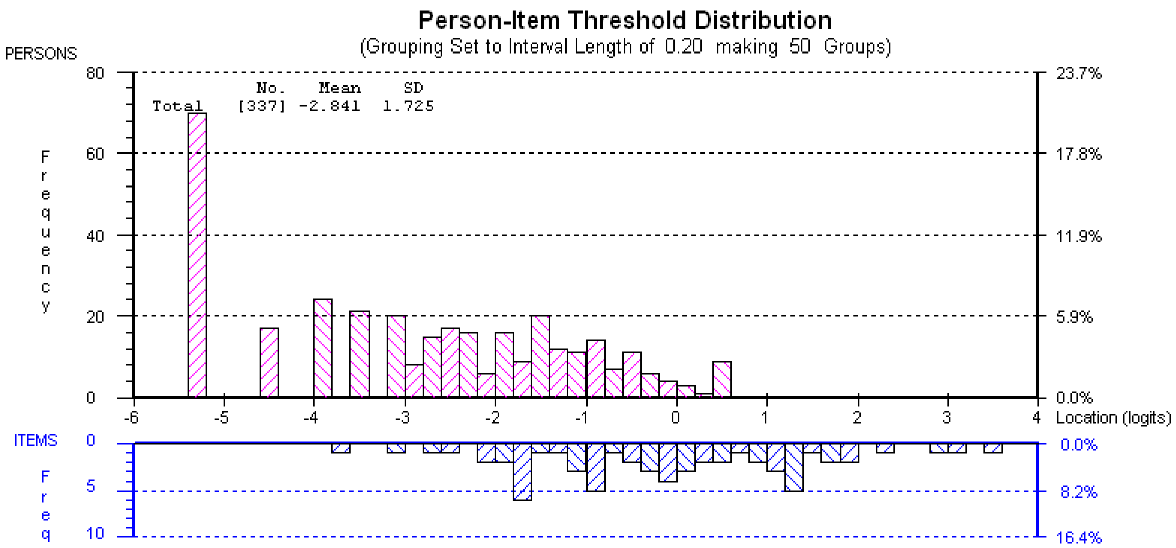
					Fit Statistics	
Subscale	**Item no.**	**Item Label**	**Location**	**SE**	**Residual**	**Chi-square**
	36	Regret of suicide attempt	0.837	0.126	-0.997	7.819
	37	Previous attempts at suicide	-0.166	0.075	**2.655**	11.564
	38	Risk of suicide	0.145	0.064	-2.477	10.028
	39	Risk of harm to self	-0.185	0.071	-0.985	9.180
	40	Future plans	0.861	0.086	0.174	7.156
Component 4 – Mediating Factors	41	Social Circumstances	0.499	0.095	0.163	8.794
	42	Protective factors	0.308	0.089	-1.884	8.142
	43	Resourcefulness	0.097	0.091	2.138	6.285
	44	Daily contact with others	0.335	0.106	-0.720	3.618
	45	Relationships	-0.754	0.099	2.429	8.873
	46	Support Networks	-0.486	0.098	**-3.416**	11.962
Component 5 – Daily Structure	47	Physical Exercise	0.467	0.068	1.654	7.155
	48	Isolation	-0.556	0.093	1.003	1.667
	49	Daily routine	0.147	0.069	-0.202	5.300
	50	Leisure Activities	0.001	0.069	-2.927	15.375
	51	Interest/Enthusiasm	-0.060	0.092	1.665	2.571
Component 6 – Risk of Harm to Others	52	Anger/agitation	-0.157	0.013	0.441	4.926
	53	Violence/hostility/aggression	-0.027	0.014	**-2.951**	34.447
	54	Risk of neglect of others	-0.003	0.014	0.461	1.995
	55	Family history of suicide	0.060	0.016	2.027	27.979
	56	Risk of harm to others	0.057	0.014	**-2.756**	19.261
	57	Domestic violence	0.083	0.017	0.231	7.517
Component 7 – Feelings/Affect	58	Overall feelings	-0.444	0.088	-0.300	11.285
	59	Tearfulness	0.276	0.094	2.360	1.020
	60	Hopelessness	0.158	0.120	-1.080	13.570
	61	Low mood/depression	0.043	0.079	-0.025	8.290
Component 8 – Basic Functioning	62	Overall appearance	-1.389	0.082	0.513	2.109
	63	General wellbeing	-0.054	0.098	-1.326	14.992
	64	Sleep	-0.363	0.084	1.400	17.651
	65	Appetite	0.179	0.088	-0.310	5.544
	66	Appropriateness of mood	0.997	0.136	-1.204	6.729

Appendix 9 – Item-Person Maps

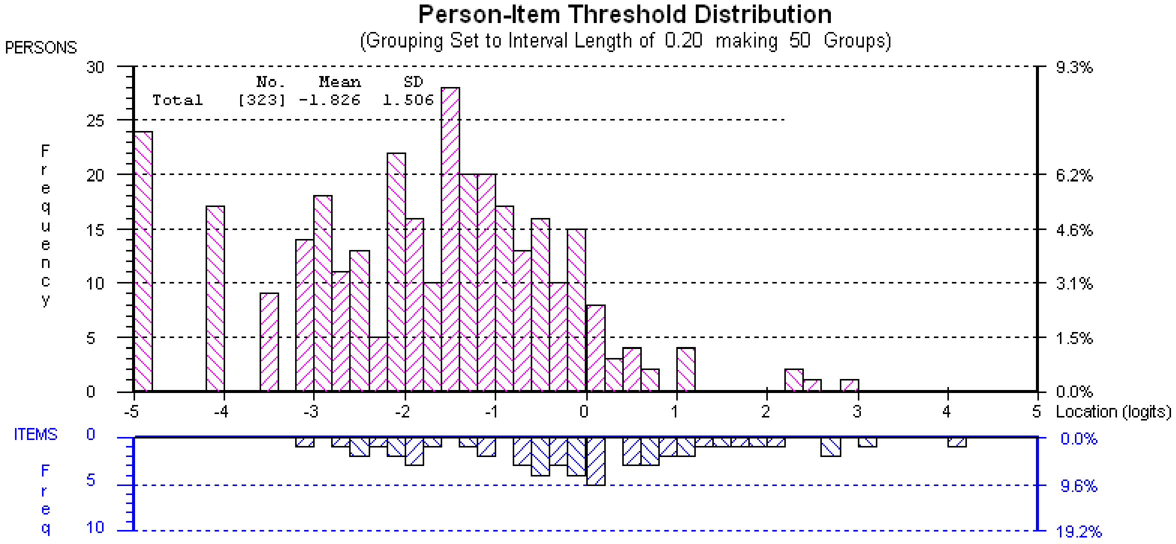
Subscale 1- Crisis Recovery Indicators



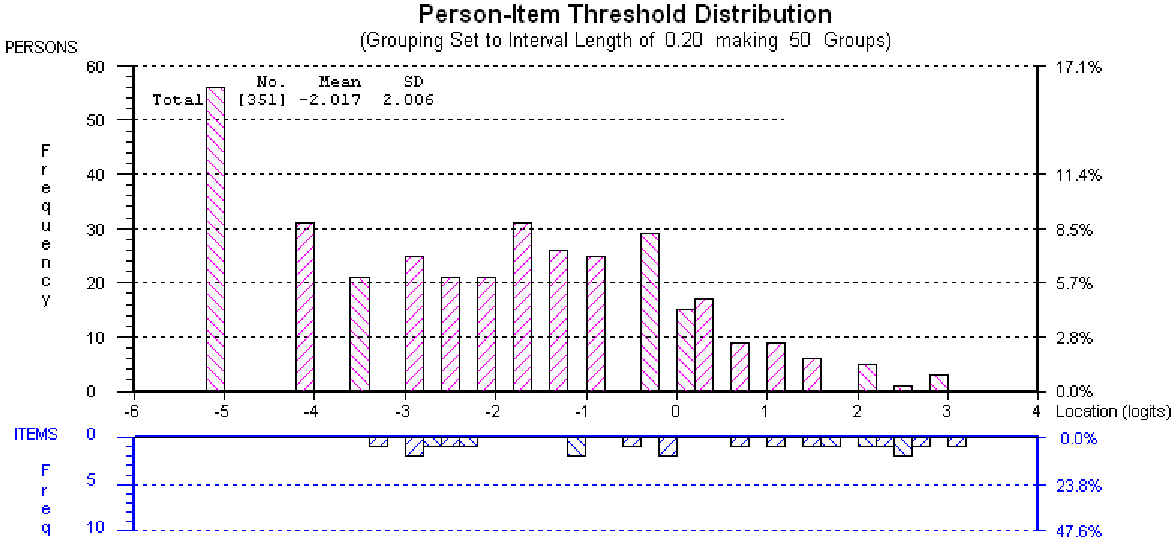
Subscale 2 – Adaptive Decision Making



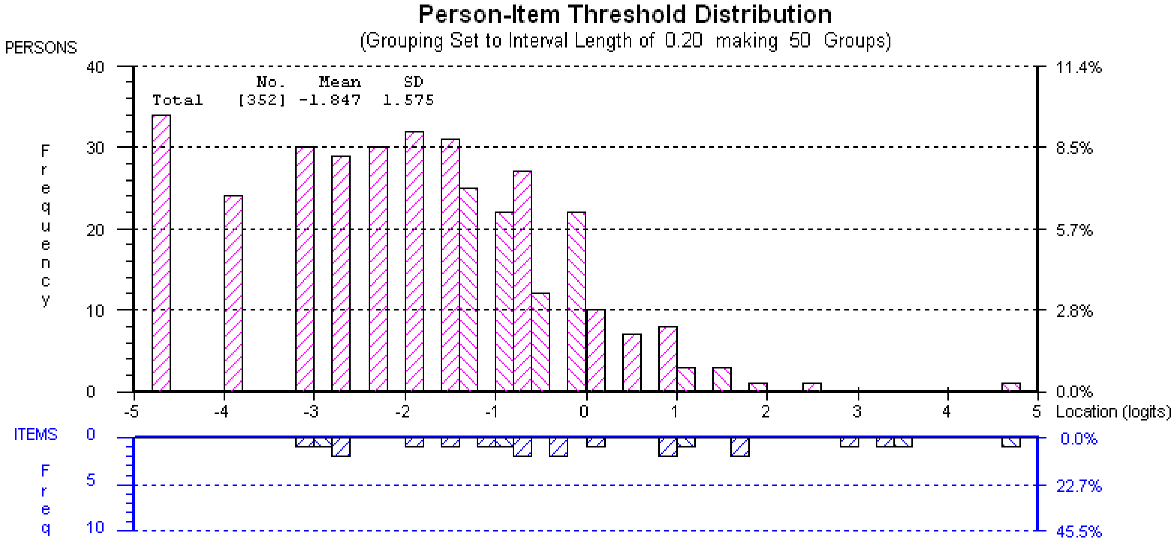
Subscale 3- Risk of harm to self



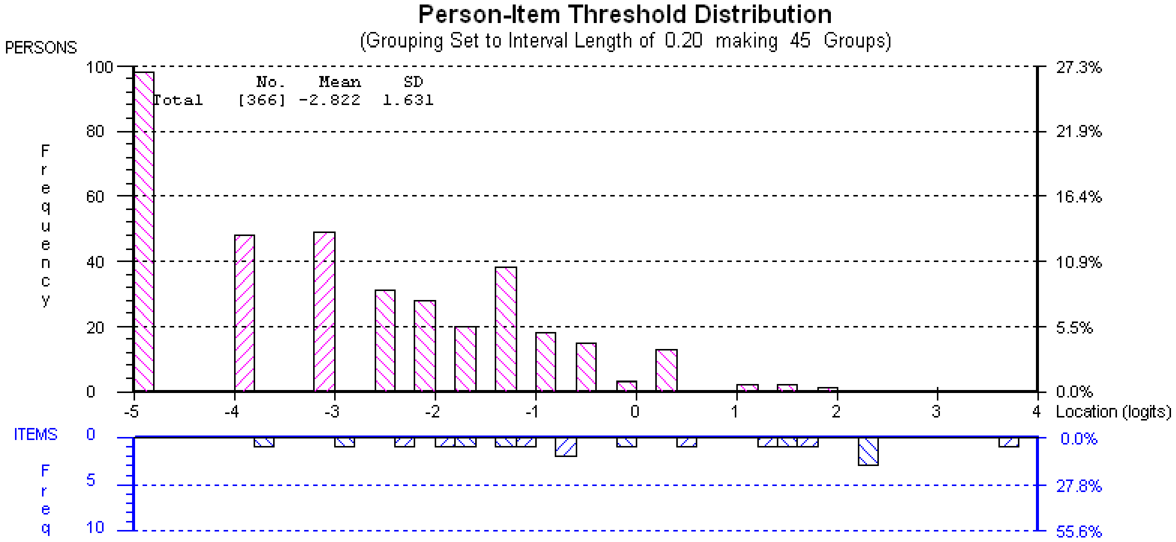
Subscale 4 – Mediating Factors



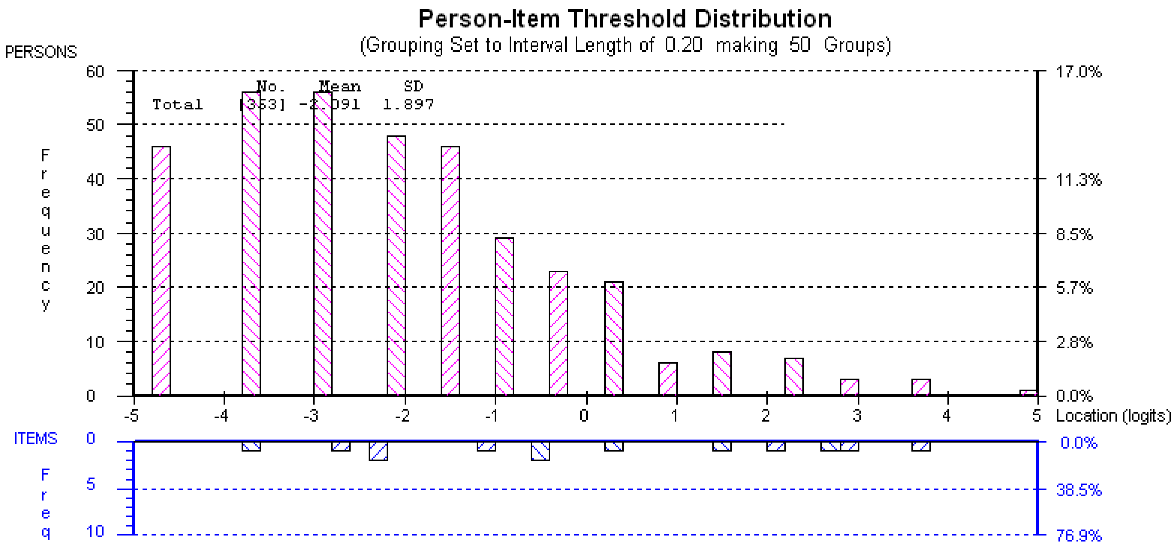
Subscale 5 – Daily Structure



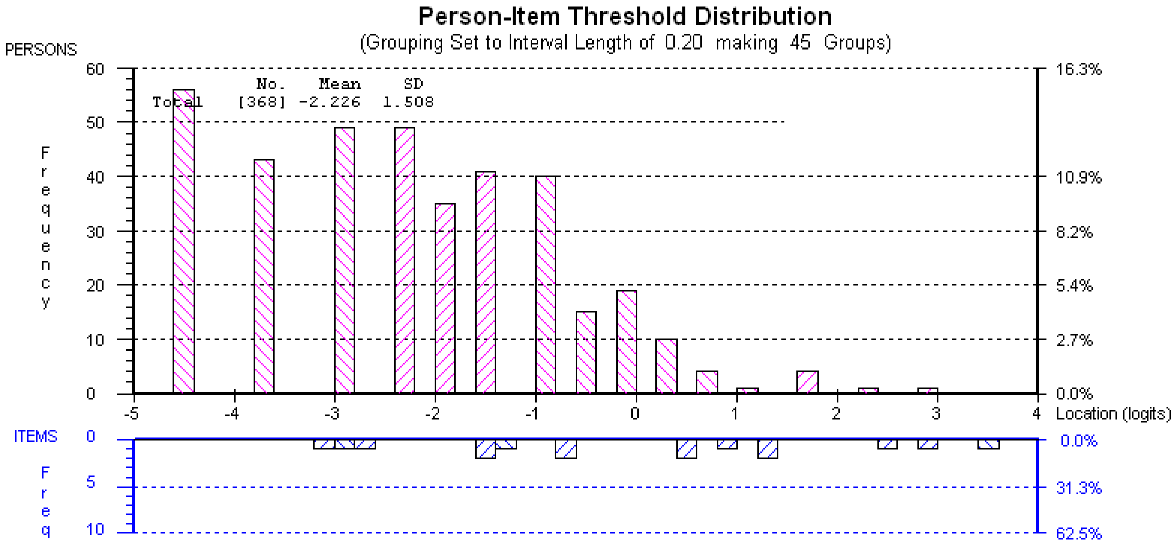
Subscale 6 – Risk of Harm to Others



Subscale 7- Feelings/Affect



Subscale8- Basic Functioning



Appendix 10 – Raw Score to Rasch Logit Scale Transformation Tables

Transformation Tables – These tables outline the 8 subscales and the percentile cut-offs based on the Rasch conversion table from the raw/ordinal level total score to logit/interval level score scale. The percentile cut-offs are calculated based on the transformed interval level Rasch logit scale and indicated in table A1 and the percentile ranks as seen by the assessor for scoring are shown in table A2.

Table A1 – Outlining the raw to Rasch transformed logit scale score for each of the subscales

Subscale 1		Subscale 2		Subscale 3		Subscale 4		Subscale 5		Subscale 6		Subscale 7		Subscale 8	
Raw	Logit	Raw	Logit	Raw	Logit	Raw	Logit	Raw	Logit	Raw	Logit	Raw	Logit	Raw	Logit
0	-6.102	0	-5.312	0	-4.924	0	-5.005	0	-4.713	0	-4.808	0	-4.789	0	-4.509
1	-5.289	1	-4.460	1	-4.099	1	-4.121	1	-3.814	1	-3.818	1	-3.715	1	-3.603
2	-4.726	2	-3.879	2	-3.541	2	-4.444	2	-3.147	2	-3.076	2	-2.860	2	-2.900
3	-4.337	3	-3.485	3	-3.164	3	-2.926	3	-2.652	3	-2.521	3	-2.180	3	-2.354
4	-4.031	4	-3.183	4	-2.874	4	-2.474	4	-2.242	4	-2.055	4	-1.563	4	-1.874
5	-3.774	5	-2.937	5	-2.631	5	-2.048	5	-1.884	5	-1.638	5	-0.969	5	-1.423
6	-3.549	6	-2.728	6	-2.418	6	-1.633	6	-1.561	6	-1.247	6	-0.384	6	-0.983
7	-3.347	7	-2.546	7	-2.225	7	-1.218	7	-1.260	7	-0.866	7	0.222	7	-0.546
8	-3.161	8	-2.384	8	-2.047	8	-0.804	8	-0.974	8	-0.484	8	0.893	8	-0.110
9	-2.988	9	-2.236	9	-1.880	9	-0.393	9	-0.694	9	-0.096	9	1.596	9	0.321
10	-2.824	10	-2.100	10	-1.723	10	0.008	10	-0.414	10	0.303	10	2.263	10	0.748
11	-2.667	11	-1.974	11	-1.576	11	0.397	11	-0.130	11	0.712	11	2.948	11	1.188
12	-2.516	12	-1.855	12	-1.438	12	0.769	12	0.164	12	1.125	12	3.783	12	1.663
13	-2.370	13	-1.743	13	-1.308	13	1.125	13	0.474	13	1.543	13	4.803	13	2.205
14	-2.227	14	-1.635	14	-1.187	14	1.465	14	0.803	14	1.974			14	2.851
15	-2.087	15	-1.532	15	-1.072	15	1.795	15	1.157	15	2.445			15	3.673
16	-1.950	16	-1.433	16	-0.965	16	2.122	16	1.543	16	3.005			16	4.720
17	-1.814	17	-1.337	17	-0.863	17	2.461	17	1.972	17	3.764				
18	-1.679	18	-1.244	18	-0.766	18	2.833	18	2.458	18	4.794				
19	-1.544	19	-1.153	19	-0.674	19	3.274	19	3.026						
20	-1.410	20	-1.064	20	-0.585	20	3.881	20	3.713						
21	-1.276	21	-0.977	21	-0.499	21	4.718	21	4.606						
22	-1.140	22	-0.891	22	-0.416			22	5.768						
23	-1.004	23	-0.807	23	-0.335										
24	-0.865	24	-0.724	24	-0.256										
25	-0.724	25	-0.642	25	-0.178										
26	-0.581	26	-0.561	26	-0.100										
27	-0.434	27	-0.480	27	-0.023										
28	-0.282	28	-0.400	28	0.055										
29	-0.127	29	-0.320	29	0.133										
30	0.034	30	-0.240	30	0.212										
31	0.199	31	-0.161	31	0.292										
32	0.369	32	-0.080	32	0.374										
33	0.542	33	0.000	33	0.459										
34	0.717	34	0.081	34	0.547										
35	0.892	35	0.163	35	0.638										
36	1.066	36	0.246	36	0.734										
37	1.237	37	0.331	37	0.834										
38	1.405	38	0.416	38	0.941										
39	1.569	39	0.504	39	1.054										
40	1.731	40	0.593	40	1.176										
41	1.891	41	0.685	41	1.309										
42	2.050	42	0.779	42	1.454										
43	2.209	43	0.876	43	1.614										
44	2.371	44	0.976	44	1.794										
45	2.536	45	1/080	45	1.998										
46	2.706	46	1.188	46	2.231										
47	2.883	47	1.300	47	2.503										
48	3.070	48	1.418	48	2.823										
49	3.269	49	1.542	49	3.212										
50	3.485	50	1.674	50	3.703										
51	3.724	51	1.813	51	4.391										
52	3.994	52	1.964	52	5.350										
53	4.313	53	2.128												
54	4.711	54	2.308												
55	5.267	55	2.511												
56	6.283	56	2.743												
57	8.138	57	3.016												
		58	3.350												
		59	3.781												
		60	4.399												
		61	5.284												

Key

20th percentile
40th percentile
60th percentile
80th percentile

Table A2 – Ranking table indicating the percentile cut-offs as seen by the assessor.

Subscale 1		Subscale 2		Subscale 3		Subscale 4		Subscale 5		Subscale 6		Subscale 7		Subscale 8	
Raw	Rank	Raw	Rank	Raw	Rank	Raw	Rank	Raw	Rank	Raw	Rank	Raw	Rank	Raw	Rank
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1		1		1		1	1	1		1	1	1		1	1
2	1	2	1	2	1	2		2	1	2		2	2	2	
3		3		3		3		3		3	2	3		3	2
4		4		4		4		4		4		4		4	
5		5		5		5	2	5		5		5		5	
6		6		6		6		6	2	6		6	3	6	
7		7		7		7		7		7		7		7	
8		8		8		8		8		8	3	8		8	3
9		9		9		9		9		9		9		9	
10		10		10		10	3	10		10		10		10	
11		11	2	11	2	11		11		11		11	4	11	
12		12		12		12		12	3	12		12		12	
13		13		13		13		13		13		13		13	4
14		14		14		14		14		14				14	
15		15		15		15		15		15	4			15	
16	2	16		16		16		16		16				16	
17		17		17		17	4	17		17					
18		18		18				18		18					
19		19		19		19		19	4						
20		20		20		20									
21		21		21		21		21							
22		22		22				22							
23		23		23											
24		24		24											
25		25		25											
26		26		26											
27		27		27											
28		28		28											
29		29		29	3										
30		30		30											
31		31		31											
32		32	3	32											
33		33		33											
34		34		34											
35		35		35											
36	3	36		36											
37		37		37											
38		38		38											
39		39		39											
40		40		40											
41		41		41											
42		42		42											
43		43		43											
44		44		44											
45		45		45											
46		46		46	4										
47		47		47											
48		48		48											
49		49		49											
50		50		50											
51	4	51													
52		52		52											
53		53	4												
54		54													
55		55													
56		56													
57		57													
		59													
		60													
		61													

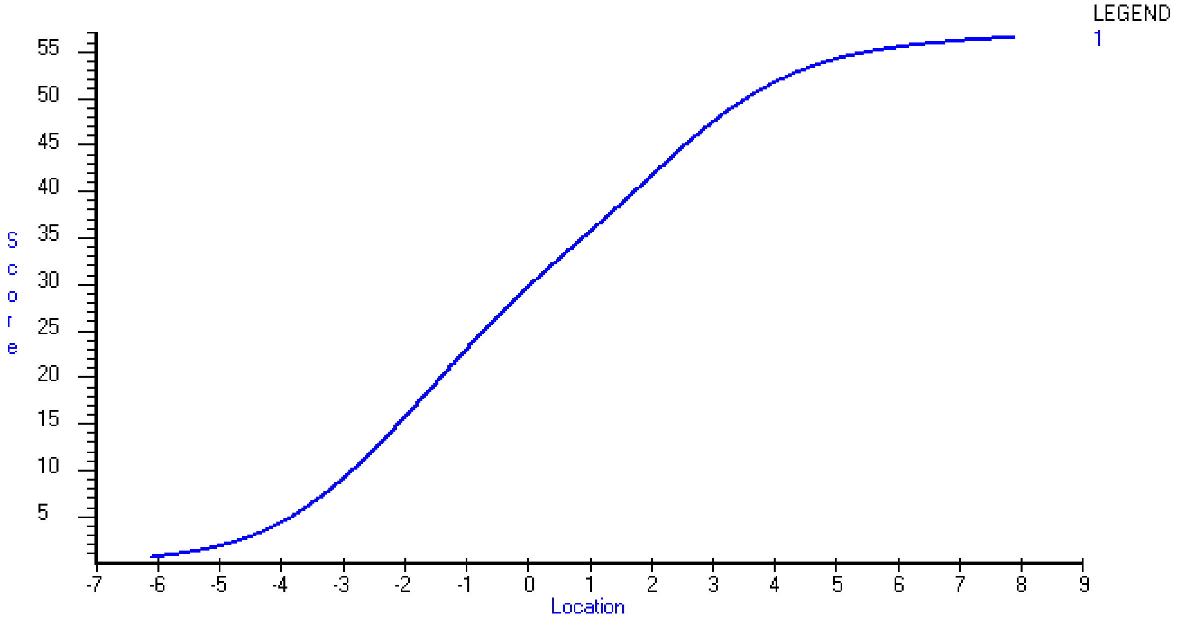
Key

1	0-20th percentile
2	20th-40th percentile
3	40th-60th percentile
4	>60th percentile

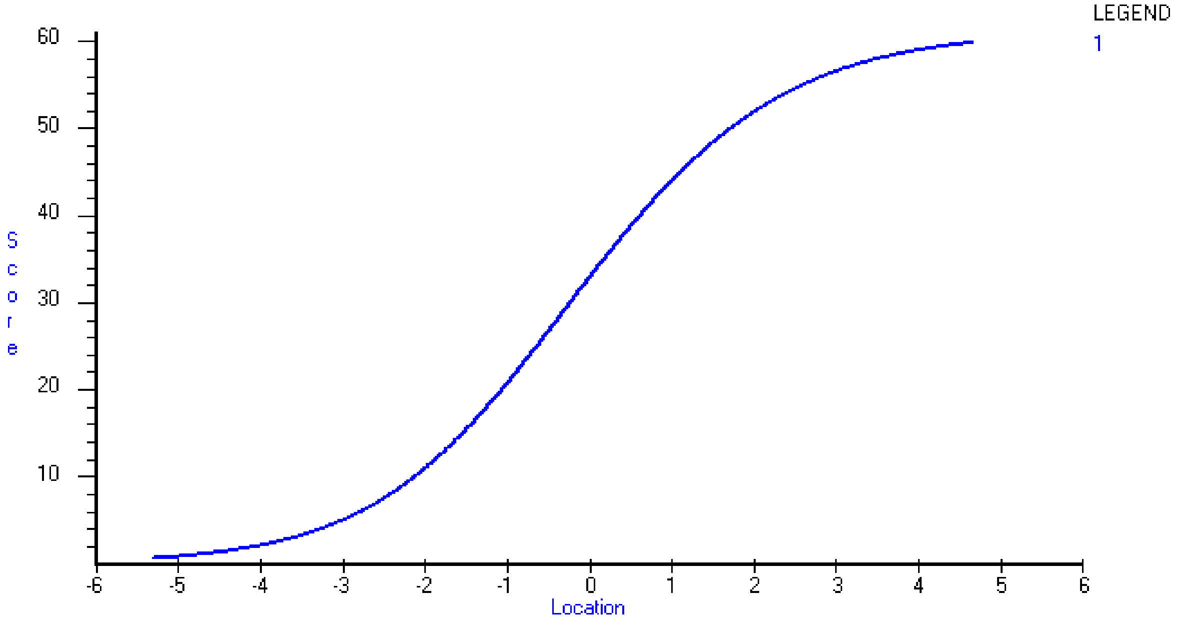
Appendix 11 – Subscale S-shape Curves

Graphs outlining the S-shaped curve created when the raw score is mapped against the Rasch logit scale for each of the measure's 8 subscales.

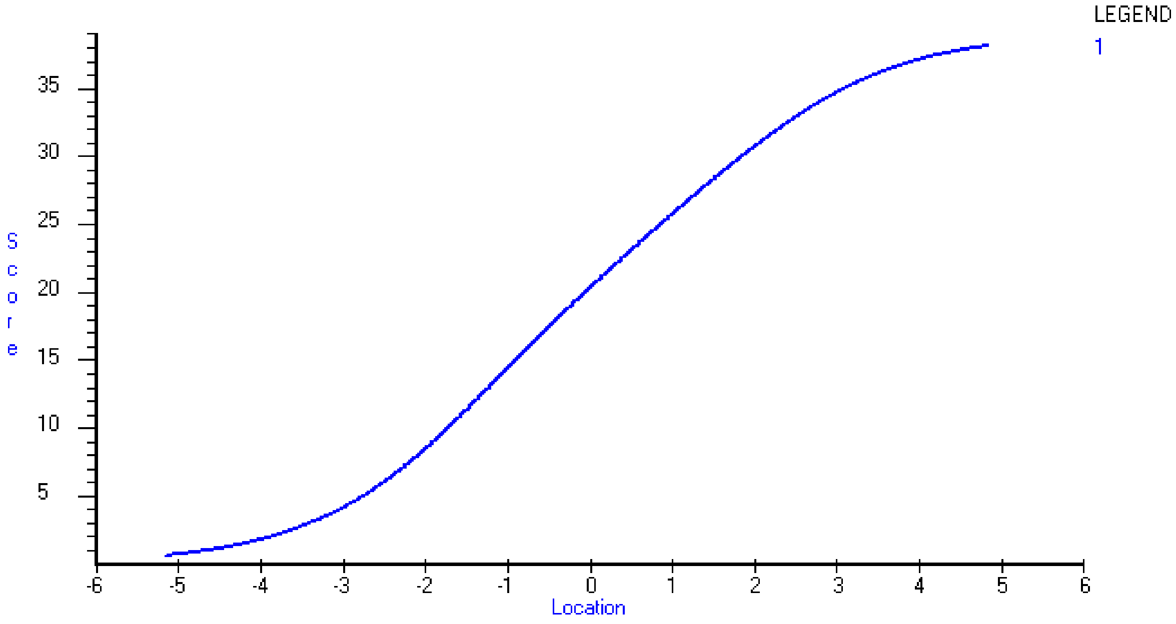
Subscale 1 – Recovery Indicators



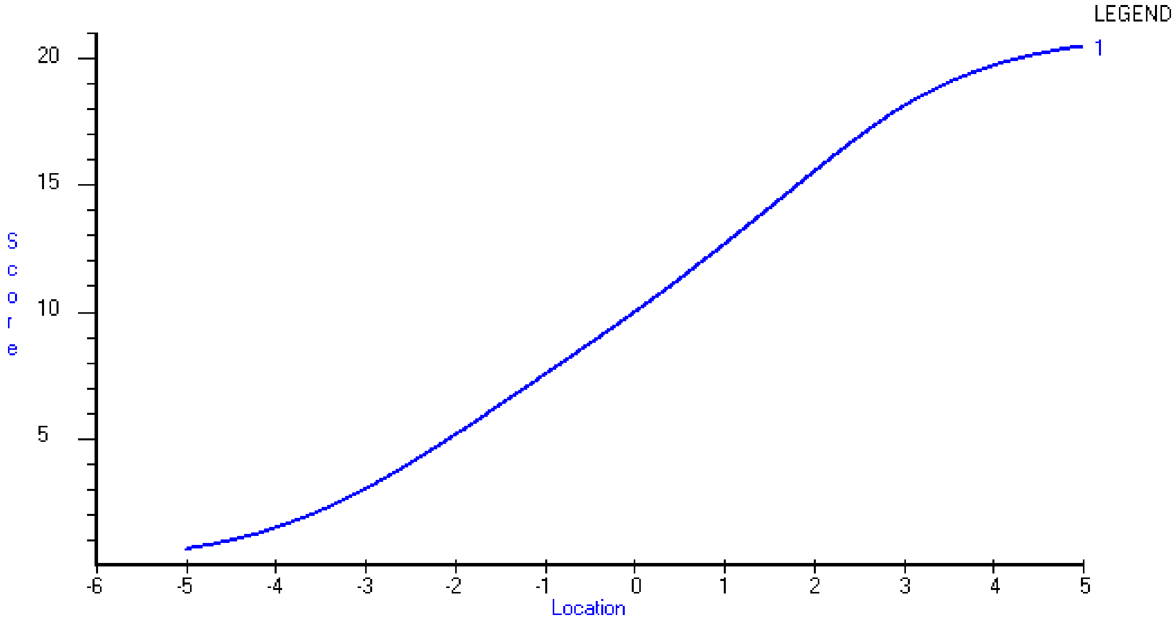
Subscale 2 – Adaptive Decision Making



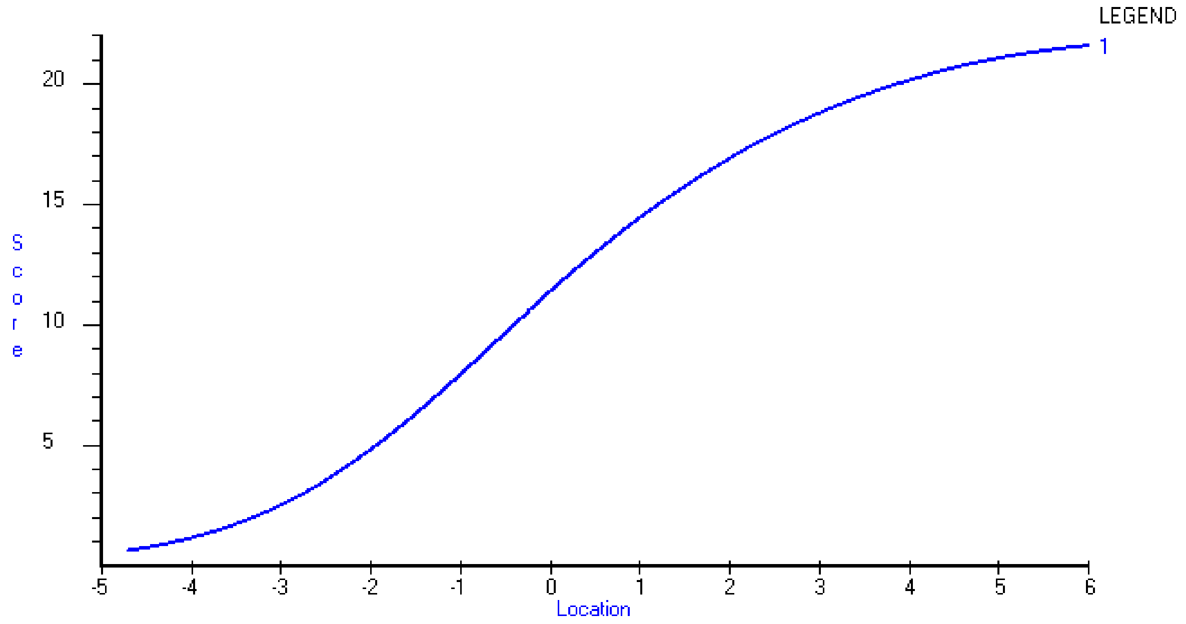
Subscale 3 – Risk of Harm to Self



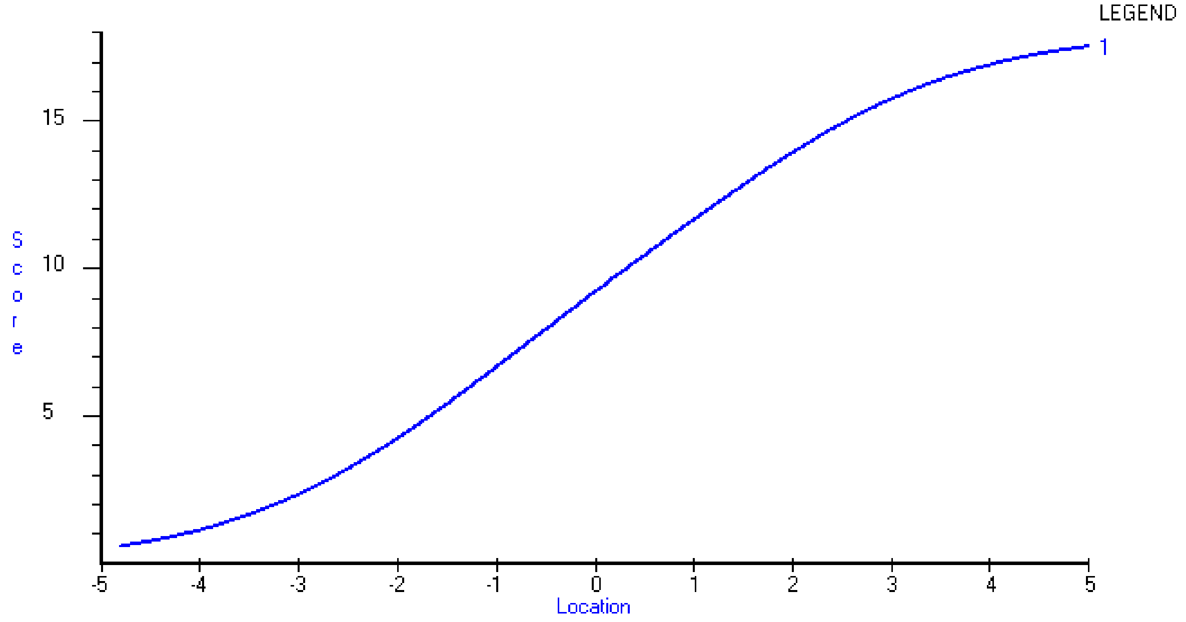
Subscale 4 – Mediating Factors



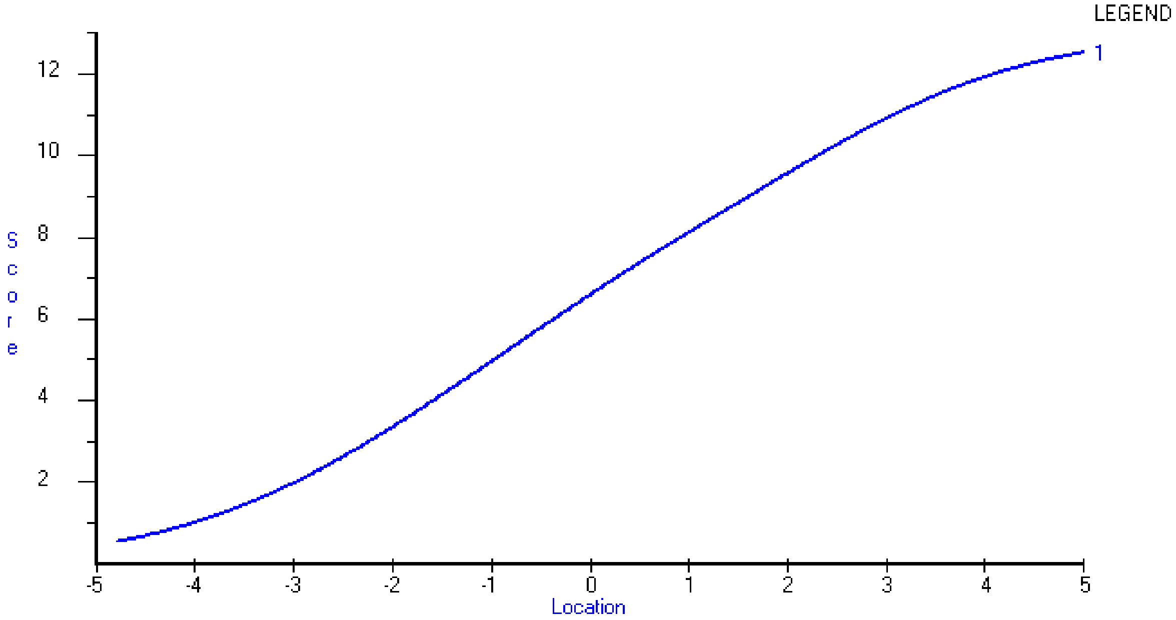
Subscale 5 – Daily Structure



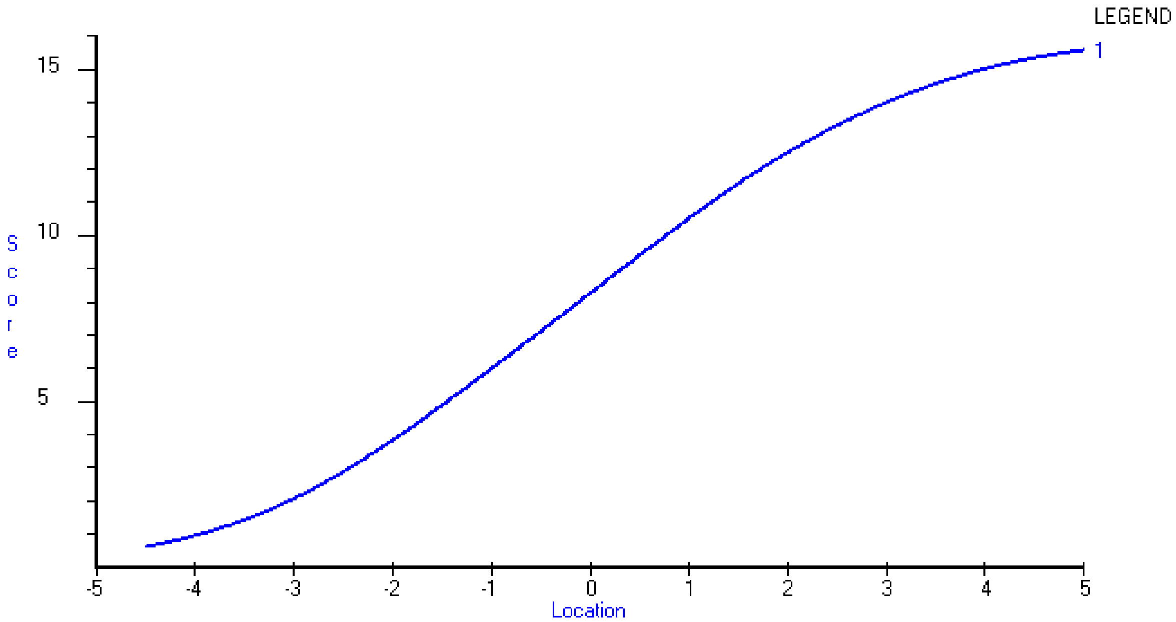
Subscale 6 – Risk of Harm to Others



Subscale 7 – Mood and Affect



Subscale 8 – Basic Needs



Appendix 12 – Item Locations and Item Fit Tables

Tables indicating the item locations, standard error and fit residuals for items comprising the 8 subscales. Item locations directly correspond to the item difficulty whilst the items' fit residual relates to the item's fit to the Rasch model. The greater the item location the greater the item difficulty. The closer the fit residual is to 0 the closer the item fits to the Rasch model and the expected outcome.

Subscale 1 – Crisis Recovery Indicators

Item no.	*Item label*	*Location*	*SE*	*Fit Residual*
1	Overall thought content and clarity	0.394	SE 0.105	-0.180
2	Concentration	1.722	0.111	0.900
3	Feelings of ineffectuality	-0.295	0.105	2.152
4	Level of need	-0.016	0.101	0.760
5	Ability to manage symptoms	-0.824	0.087	0.631
6	Acceptance of difficulties	-0.006	0.109	-0.221
7	Ability to relax	0.389	0.139	0.510
8	Stability of presentation	-0.119	0.097	-0.372
9	Staff Intuition	-0.589	0.104	1.278
10	Level of functioning	0.014	0.110	-1.213
11	Energy/get up and go	-0.513	0.105	0.265
12	Change from normal presentation	0.085	0.089	-1.015
13	Predictability	-0.393	0.105	-0.723
14	Intensity of symptoms	-0.471	0.082	-2.538
15	Responsibility for self	-0.240	0.105	0.381
16	Ability to take control	0.862	0.105	0.417

Subscale 2 – Adaptive Decision Making

Item no.	*Item label*	*Location*	*SE*	*Fit Residual*
17	Thought block	-0.201	0.102	-1.235
18	Stream of thought	0.166	0.097	-0.788
19	Flight of ideas	0.628	0.127	-1.180
20	Poverty of thought	0.651	0.112	-0.676
21	Understanding of reality	-0.588	0.063	3.036
22	Capacity to consent	0.126	0.112	-3.234
23	Judgement	-0.875	0.055	2.810
24	Confusion	0.127	0.096	0.466
25	Insight	-0.939	0.072	1.092
26	Irrational speech	0.617	0.116	-2.492
27	Overall acceptance of support	0.488	0.120	0.069
28	Ability to rationalise	-0.773	0.072	1.894
29	Speech	0.687	0.094	-0.043
30	Response to Hallucinations/Delusions	-0.074	0.091	1.253

Subscale 3 – Risk of harm to self

Item no.	*Item label*	*Location*	*SE*	*Fit Residual*
31	Impulsivity	-0.536	0.066	1.356
32	Regret of actions during crisis	0.015	0.063	2.265
33	Overall risk	-0.964	0.053	-0.972
34	Access to lethal means	-0.243	0.060	-0.426
35	Intent to commit suicide	0.237	0.066	-2.103
36	Regret of suicide attempt	0.837	0.126	-0.997
37	Previous attempts at suicide	-0.166	0.075	2.655
38	Risk of suicide	0.145	0.064	-2.477
39	Risk of harm to self	-0.185	0.071	-0.985
40	Future plans	0.861	0.086	0.174

Subscale 4 – Mediating Factors

Item no.	*Item label*	*Location*	*SE*	*Fit Residual*
41	Social Circumstances	0.499	0.095	0.163
42	Protective factors	0.308	0.089	-1.884
43	Resourcefulness	0.097	0.091	2.138
44	Daily contact with others	0.335	0.106	-0.720
45	Relationships	-0.754	0.099	2.429
46	Support Networks	-0.486	0.098	-3.416

Subscale 5 – Daily routine

Item no.	*Item label*	*Location*	*SE*	*Fit Residual*
47	Physical exercise	0.467	0.068	1.654
48	Isolation	-0.556	0.093	1.003
49	Daily routine	0.147	0.069	-0.202
50	Leisure activity	0.001	0.069	-2.927
51	Interest	-0.060	0.092	1.665

Subscale 6 – Risk of Harm to Self

Item no.	*Item label*	*Location*	*SE*	*Fit Residual*
52	Anger/agitation	-1.179	0.100	0.535
53	Violence/hostility/aggression	-0.269	0.092	-2.831
54	Risk of neglect of others	-0.340	0.100	1.622
55	Family history of suicide	0.092	0.133	1.893
56	Risk of harm to others	0.377	0.101	-3.074
57	Domestic violence	0.906	0.141	1.394

Subscale 7 – Mood/Affect

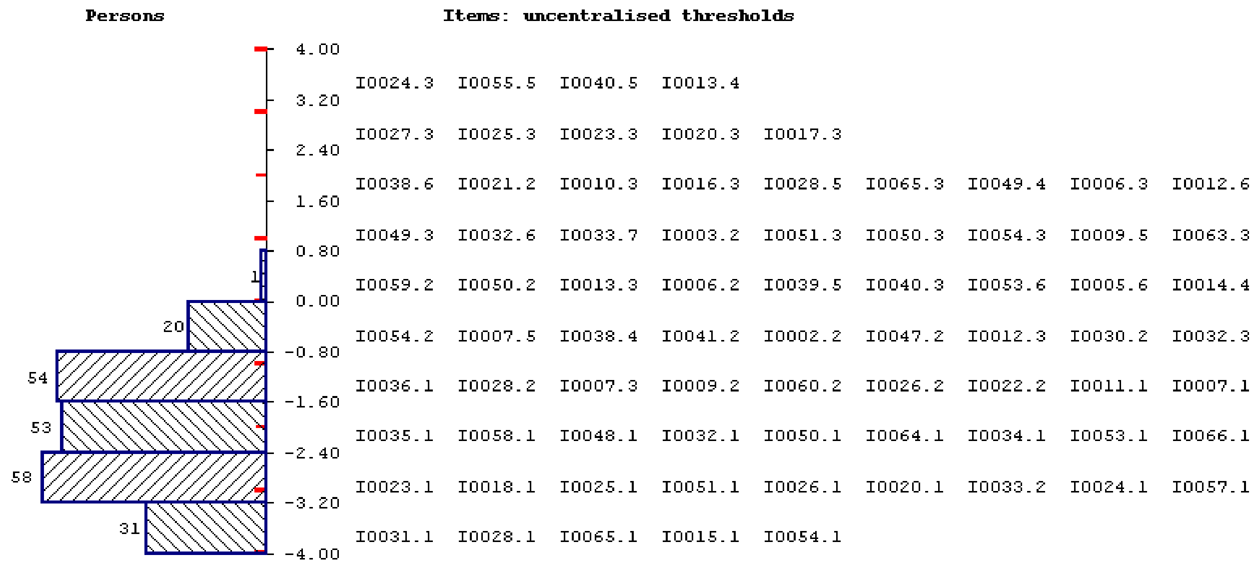
Item no.	*Item label*	*Location*	*SE*	*Fit Residual*
58	Overall feelings	-0.444	0.088	-0.300
59	Tearfulness	0.276	0.094	2.360
60	Hopelessness	0.158	0.120	-1.080
61	Low mood/depression	0.043	0.079	-0.025

Subscale 8 – Basic Needs

Item no.	*Item label*	*Location*	*SE*	*Fit Residual*
62	Overall appearance	-1.389	0.082	0.513
63	General wellbeing	-0.054	0.098	-1.362
64	Sleep	-0.363	0.084	1.400
65	Appetite	0.179	0.088	-0.310
66	Appropriateness of mood	0.997	0.136	-1.204

Appendix 13 – Item-Person Location Map

This is a sample section of the item category-person map to provide an example of the item-person location map. The numbers indicate the original item numbering of the 143 item measure and due to the set-up of the RUMM2030 software it was not possible to extract the full Figure in the preferred standard format. The numbers before the decimal point indicate the item number and the number after the decimal point indicates the rating scale category. For example, according to this item-person location map the least likely item to receive a rating is item 13 on the 4th category of the item rating scale. Item 13 is identified as the item 'Speech'.



Appendix 14 – Definition List for the 66 Item CRAFT

1.0	**Recovery Indicators**	
1.1	Overall thought content and clarity	The perceptions, convictions, occupation and impulses in an individual's thoughts. The accessibility and ability to understand clearly one's own thoughts. Does the individual have difficulties in this area? Are able to manage it well? Has this improved or deteriorated since experiencing crisis? Does their thought content and clarity cause concern in relation to treating the individual at home?
1.2	Concentration/attention/memory	The individual's ability to focus attention for a sustained period of time on a specific stimulus, sensation, idea, thought or activity. Has the individual's concentration/attention/memory been affected in anyway by the crisis? Is this change a cause for concern? Is the change in their concentration/attention/memory having an impact on their ability to function? Could this put them at risk for example if they put something on the stove to cook but then forgot about it?
1.3	Feelings of ineffectuality	The feeling that one's own actions will be fruitless and pointless with an inability to influence events in their own life. Are their feelings of ineffectuality a new phenomenon or a long standing issue? Does it have an impact on their ability to live a functional life? Are their feelings of ineffectuality a cause for concern to treat this person at home? Will they be able to successfully care for themselves? Do they have a suitable support network?
1.4	Level of need/dependence	This item looks at how dependent on others the individual is. Are they able to cope on their own at all or do they constantly look to others to cope for them? Do they need other people around them to manage their difficulties? How is this impacting on those around them? Is their level of need or level of dependence a cause for concern?
1.5	Ability to manage symptoms	This is looking at the capacity of a person to cope with their mental health difficulties. Are they able to manage their symptoms or do they find them overwhelming? Are they able to employ helpful coping strategies to manage difficulties? Is the person's ability to manage their symptoms a cause for concern?
1.6	Acceptance of difficulties	This is a person's ability to understand their difficulties or parts of their life. This can be a healthy response to difficulties that cannot be changed such as permanent physical problems like body shape or a physical disability. However, it can also be an unhelpful response when a person accepts difficulties that are within their power to change. Is this person's acceptance of their difficulties a cause for concern or a helpful coping strategy?
1.7	Ability to relax	The ability to rest the body and mind. Is this person able to participate in activities for relaxation? Has their ability to relax changed since experiencing crisis? Are they able to be at peace in anyway e.g. sitting reading a book, spending time in the garden, engaging in relaxation exercises etc.?
1.8	Stability of presentation	This is the constancy and continuity of the individual's presentation. Is their presentation very changeable and unpredictable? Does their presentation make it difficult to plan their care? Does the changeability of their presentation make it difficult to treat them at home? Is the stability of their presentation a cause for concern?

1.9	Staff intuition/Instinct	This is defined as the immediate understanding, knowledge, or awareness, derived neither from perception nor from reasoning. This is asking you to assess your own instinctual reactions/gut instincts to the situation and the Service User's presentation. There may be times when the Service User presents as 'well' but the assessor has a feeling (an instinct), that not all is as it appears. This can be shown as having an uneasy or uncomfortable feeling in response to the situation. As a definition this is having an understanding of the truth that is not based on any facts or evidence. Are these feelings a cause for concern?
1.10	Level of functioning	This is the ability of the person to carry out the basic and expected actions of an adult. Has their level of functioning changed? Is this a cause for concern?
1.11	Energy/Get up and go/Drive	This is looking to assess a person's ambition and enthusiasm for life.
1.12	Change from normal presentation	This is assesses the differences observed from a person's normal presentation. How much of a change has occurred? Is this change a positive or negative change? Is this change a cause for concern?
1.13	Predictability	The degree to which the Service User's future behaviour/presentation can be estimated. Is this person known well enough to be able to predict their future actions? Are they new to the team and therefore unpredictable? Would you be able to predict their actions over the next few days/weeks/months with any accuracy? Is this Service User's predictability a cause for concern for treating this person at home?
1.14	Intensity of symptoms	This is the strength, concentration and magnitude of presenting symptoms. Is the intensity of the symptoms a cause for concern?
1.15	Responsibility for self/Independence	This is the ability of the Service User to take responsibility for their own actions and to be independent. Has this person's level of independence changed since going into crisis? Is this a cause for concern for this individual to manage at home?
1.16	Ability to take control	This is the person's ability to manage situations effectively and to steer situations towards desired outcomes. Is this person able to be assertive and proactive? Are they passive in nature? Is their ability to take control supporting them in recovering from the crisis or is it a hindrance? Is their ability to take control a cause for concern?
2.0	**Adaptive Decision Making**	
2.1	Thought block	This is the interruption or obstruction to the train of speech/thought. Does the individual stop mid-sentence? When asked a question does it appear that they are unable to answer? To what extent? Will this interfere in their life or affect their ability to function, to have their needs met, to deal with difficulties? How much of a cause for concern is this difficulty?
2.2	Stream of thought	This looks at the continuity, coherence, content, preoccupation, amount and productivity of thought, normally assessed through conversation. Is the individual able to follow a topic of conversation or do they go off on unrelated tangents? Do they make sense? Is their thinking too fast or too slow? Has this changed as a result of crisis? Are they able to cope with their stream of thought or is it causing them difficulties? Is their stream of thought a cause for concern?
2.3	Flight of ideas	The experience of having lots of thoughts leading to ideas that race from topic to topic but generally associated. This can make it difficult to hold a two way conversation or to complete a conversation before starting on the

		next one. Does their flight of ideas pose a cause for concern? Does it interfere with their ability to communicate with others, to function? Are they able to manage their flight of ideas?
2.4	Poverty of thought (speech)	The form of thought is disordered and this is often displayed in their speech. Is this a cause for concern for the individual? Are they able to communicate their needs? Does this pose a risk to their wellbeing or the wellbeing of others? Has their thinking changed since experiencing crisis? Is poverty of thought normal for this person's presentation?
2.5	Understanding of reality	This is the ability of the individual to perceive the world in the realm of what would be considered normal. Are they able to perceive the world around them as most people do? Are they able to understand the cause/effect links as others do? Is their understanding of reality alarming? Is it a cause for concern? Has it changed in any way since experiencing crisis or is this understanding normal for them?
2.6	Capacity to consent	The ability to receive, contain and weigh up information that will support a person to make an informed decision to either agree or disagree with a suggestion. The person also needs to have the ability to communicate that decision. Does this person have capacity to consent? To what extent? Is their capacity to consent a cause for concern? Is there someone who is able to support them in making decisions and protect their interests? Has their capacity to consent changed since experiencing crisis?
2.7	Judgement	The ability to respond to situations using knowledge of what is normal, normal customs and expectations of society. The ability to form an opinion from circumstances presented in a rational and logical fashion that is congruent with the culture of that person. Has the individual's judgement been changed or impaired in any way? Will their judgement be a cause for concern for treating this person at home? Is their judgement a cause for concern in all situations or does depend on the situation?
2.8	Confusion	This is the feeling of being unsure or unclear. Has their level of confusion changed since experiencing crisis or is this normal for them? Is it due to another mental health or organic problem? Does their level of confusion make them vulnerable or put them/others at risk? Can they be safely treated at home or is this a cause for concern?
2.9	Insight	This is defined as a clear and deep understanding or perception. The degree to which the Service User acknowledges and comprehends his or her mental disorder and its effect on others. Does this person have insight into their difficulties? Are they able to use this insight to manage their problems better? Is this a cause for concern for home treatment?
2.10	Irrational speech	This is when a person is deprived of normal mental clarity or sound judgment. They will talk in a way that is not in accordance with reason. Is this person talking in an irrational way? Is this a cause for concern for treating the person at home.
2.11	Overall acceptance of support	An individual's overall openness to support from others. Is this person able to accept support from others or do they chose not to? Do they allow friends and family to help them when they need it? Is their ability to accept support a cause for concern?
2.12	Ability to rationalise	This is the ability to employ reason and to work through problems in a logical manner. Is this person able to be rational about their difficulties? Has their ability to rationalise been affected by their current crisis? Is their ability to rationalise a cause for concern?

2.13	Speech	This is the ability of the Service User to communicate verbally. Are they able to communicate their needs to others? Is the Service Users' ability to communicate a cause for concern?
2.14	Response to delusions/hallucinations/voices	This is looking at how the person responds to the experience of delusions and hallucinations. Are they able to manage this experience well, being able to decide what is real and what is not and cope appropriately? Do they act on their hallucinations? Is their response to delusions/hallucinations/voices a cause for concern?
3.0	**Risk of Harm to Self**	
3.1	Impulsivity	How much of a cause for concern is the individual's inclination to act on impulse rather than thinking things through? Are they able to manage their impulsive urges, drives or temptations to behave in a way that would be unhelpful or damaging to themselves or others? Has their level of impulsivity changed as a result of the crisis?
3.2	Regret of actions during crisis	This is the feeling of disappointment about their behaviour and actions during the crisis period? Is this feeling appropriate? Is this a healthy reaction to their crisis? Is this a cause for concern?
3.3	Overall risk	This is trying to assess the overall risk of the Service User in terms of harm or neglect to self or others. Are they likely to expose themselves or others to loss or injury? How predictable is this person? Are the possible outcomes for this person in their current situation balanced towards the positive or negative? Is this person's overall risk presentation a cause for concern? Are the risks manageable? Is there enough support in place to manage this risk in the community?
3.4	Access to lethal means	This is when a person has access to a deadly or fatal method capable of causing death. It is not simply having access to lethal means in itself that is being assessed here, as it can be assumed that anyone has access to lethal means through kitchen utensils, over the counter medication etc. What is being assessed here is the cause for concern that the person has access to lethal means and intends to use that access for detrimental purposes. For example are they hoarding medication? Have they purchased materials for the purpose of taking their own life?
3.5	Intent to commit suicide	This is the Service User's determination to purposefully end their own life in the future. Have they made a plan? Have they put their affairs in order? A person may try and take their own life in a manner that, as professionals, we know will not be effective (e.g. a large overdose of certain homeopathic medicines), however the intent to die may be very high and would still be a cause for concern. Is this person intent on taking their own life? If they have failed this time will they persist?
3.6	Regret of suicide attempt	Does the Service User regret their attempt at suicide indicating that they wish to carry on and to work through the crisis? Or do they regret that the suicide attempt didn't work? How much of a cause for concern is this individual's reaction to their attempted suicide? If the Service User has not made an attempt to commit suicide this counts as a 'not cause for concern' and would be considered a protective factor.
3.7	Previous attempts at suicide	The knowledge that the SU has made previous attempts to take their own life. How serious were these attempts? Assessment of intent, planning, precautions the person took to not be found, seeking help afterwards, method, and final acts such as settling affairs, how recent the attempt was etc. need to be assessed to rate this item. Should the previous attempts be a

		cause for concern for supporting the Service User at home?
3.8	Risk of suicide	What is the overall risk of the Service User purposefully taking their own life? Is this risk manageable by the team to enable home treatment? Risk factors include age, sex, psychiatric disorder, previous attempts at suicide, social isolation, unemployment, marital status, profession/social class. Other risk factors include chronic painful physical conditions, debilitating neurological disorders, unresolved current problems especially acute single events.
3.9	Risk of harm to self	This is the threat that the person will harm themselves. Is there any evidence of self-harm? Have they expressed any thoughts to harm themselves? Are they able to manage/challenge these thoughts or are they wishing to carry them out?
3.10	Future plans	The type of future goals and plans a Service User has made. How realistic and achievable are these plans? Do they have support to achieve their goals and to manage any setbacks? Have they made clear plans or is it vague? Has there been a change in their future goals since experiencing the crisis? Is this a cause for concern?
4.0	**Mediating Factors**	
4.1	Social circumstances	This includes all the structural factors that allow a person to live their life for example housing, money, welfare etc. Are their social circumstances causing stress in the person's life? Are they able to manage this stress or is it becoming a burden?
4.2	Protective factors	The factors in the individual's life that will prevent them from carrying out a harmful act on themselves or others or prevents them from neglecting themselves or others. Do they have anything in their life which makes them want to carry on trying and wishing to make positive changes in their life?
4.3	Resourcefulness	This is the ability to deal skilfully and promptly with new situations and difficulties. This requires that the person is able to pull on their own strengths and resources as well as the resources around them. How well is this person able to utilise the resources available to them? Does this help them to manage their crisis more easily? Is their resourcefulness a cause for concern?
4.4	Daily contact with others	Does the Service User see other people every day? Do they live on their own or do they live with others? During a crisis period an individual's presentation can change quickly and therefore it can be important to know that there are other people around (other than the CRHT team) who will be able to keep an eye on that person's progress. Is it important for this Service User? Is their daily contact with others causing you concern?
4.5	Relationships	This looks at the connections, relationships and associations that the Service User has with other people. Is the Service User able to maintain healthy relationships? Has there been a change in the relationships they hold since experiencing crisis? Are the relationships they hold helpful or unhelpful? Would these relationships be a cause for concern?
4.6	Support networks	These are the different relationships with people and groups that support a person to function both physically and emotionally. Does this person have any support networks? Are they able to tap into these networks when necessary?

5.0	**Daily Structure**	
5.1	Physical exercise	How much of a cause for concern is their level physical exercise? Do they have a healthy approach to exercise? Have their levels of exercise changed dramatically since experiencing crisis for example exercising too much or a marked reduction in exercise?
5.2	Isolation	This can mean an actual *physica*l separation of a person from others *or* the *feeling* of being disliked or alone. Is isolation normal for this person? Have they recently started isolating themselves due to mental health reasons? Is this an unhelpful/helpful coping strategy? Is their level of isolation concerning?
5.3	Daily routine	In basic terms this is the order of events that a person regularly follows each day. Has their daily routine changed in any way since experiencing crisis? If so, has this been an improvement or deterioration? It focuses on the order in which activities are done e.g. a person’s routine may be disturbed so that they are getting up much later or performing tasks out of order or staying up all night.
5.4	Leisure activities/Hobbies	Time free from the demands of work or duty, when one can rest, enjoy hobbies or sports. Is the individual still able to enjoy their free time? Are they participating in their normal leisure activities and hobbies? Has this changed in anyway over the course of the crisis?
5.5	Interest/Enthusiasm	This is when an individual’s attention, concern or curiosity is particularly engaged by something. Is this person still showing an interest/enthusiasm in the areas of their life that they used to? Or to new areas? Have their levels of enthusiasm/interest changed? Has this been a negative or a positive change? Is it a cause for concern?
6.0	**Risk of Harm to Others**	
6.1	Anger/Agitation	This is when a person has a strong feeling of displeasure and an emotional state of restlessness. Is the Service User showing any signs of anger/agitation? Is this an appropriate emotional response in the given situation or is this reaction a cause for concern?
6.2	Violence/Aggression/Hostility	This is when a person asserts a rough or injurious physical force, action or treatment. This can be in the form of making assaults or attacks. Are this person’s levels of violence/aggression/hostility a cause for concern for treating them at home? This focuses on violence by the Service User towards others. For violence towards the Service User please use item 103 ‘Overall Vulnerability’.
6.3	Risk of neglect of others	This is when a person shows lack of care or poor treatment of others. This is often in relation to dependents such as children, the elderly or the disabled but may also relate to peers. Is this person neglecting others or are they able to offer the care and treatment that needed? Is their neglect of others a cause for concern? Is there anyone else who can take over in these caring roles for them while they resolve their crisis?
6.4	Family history of suicide	The knowledge that other members of the Service User's family have made suicide attempts. Has this had an impact on the Service User? Is this a cause for concern? Family history of suicide has previously been shown in research to be a risk indicator for completed suicide.
6.5	Risk of harm to others	This is the threat that the person will harm others. How real is this threat to

		harm others? Can this person be treated at home or do they pose to much of a risk? How much of a cause for concern is this person?
6.6	Risk of domestic violence	This is when a family member, partner or ex-partner attempts to physically or psychologically dominate the other. Is this person at risk of domestic violence? Is domestic violence a cause for concern?
7.0	**Feelings and Affect**	
7.1	Overall feelings/Mood state	This is their overall emotional state. Mood is defined as a temporary but relatively sustained and pervasive affective state, often referred to as emotion. Does the individual's mood change quickly and/or often? Has their mood changed significantly due to their crisis? Are they able to manage their mood/feelings/emotions or do they find it difficult to cope with? Are their feelings/mood in context to the current situation? Are their overall feelings or mood state a cause for concern?
7.2	Tearfulness	This is when an individual is easily brought to tears in response to situations that would not normally warrant this reaction. Is this normal for this person? Are they tearful in all situations or does it vary? Have they become more or less tearful recently? Is this behaviour a cause for concern in this person?
7.3	Hopelessness	This is when the individual has the feeling that conditions will never improve and that there is no solution to a problem. The experience of complete hopelessness is one of the strongest indicators of intent to commit suicide and can indicate that an individual has completely given up and does not see the point in trying to make changes. Is this individual feeling hopeless? Is it a cause for concern in this person or are they able to manage this feeling? Is there adequate support in place to allow this person to be treated at home?
7.4	Low mood/Depression	Low mood or depression is the feeling of sadness, being gloomy, downcast and experiencing emotional dejection. When a person experiences extreme low mood/depression they often find it difficult to view anything in either a neutral or positive way, tending to only look at the negatives. It is defined as a mood state of sadness, gloom, and pessimistic ideation, with loss of interest or pleasure in normally enjoyable activities, accompanied in severe cases by weight loss, feelings of worthlessness and guilt, diminished ability to think or concentrate and recurrent thoughts of death or suicide. Is this Service User able to look for the positives? Are they able to manage their low mood or is it having an impact on their ability to function? Is it having a detrimental effect on those around them? Does their low mood cause concern for treating this person at home?
8.0	**Basic Needs**	
8.1	Overall appearance	This is looking at the person's overall appearance including the way they are dressed, how well groomed they are, personal hygiene, whether they look healthy in themselves e.g. weight, skin tone etc. Is their overall appearance a cause for concern?
8.2	General wellbeing	This is the Service User's overall general state of physical and psychological health. How is their sleep? Their appetite? Has their weight changed? Are they as energetic and full of life as before the crisis or has this changed? Is there overall general wellbeing a cause for concern? Have they been ill more regularly recently?
8.3	Sleep	The ability of a Service User to get the quality and quantity of sleep necessary for normal functioning. Have there been any changes in their

		sleep as a result of the crisis? Are they sleeping much more or much less? Are they getting quality sleep or do they still feel tired the next day? Is their sleep a cause for concern or is this normal for them?
8.4	Appetite	This is the person's instinctive physical desire for food and/or drink. Has their appetite changed in anyway due to the crisis? Is it a cause for concern? For example if they have lost their appetite are they getting enough nutrition to be healthy?
8.5	Appropriateness of mood	Is the individual's mood in context? Is it an appropriate mood in the situation/circumstances? Would others react in a similar way? Is the appropriateness of their mood a cause for concern?

Appendix 15 – CRAFT Scoring Templates

Subscale 1 – Crisis Recovery Indicators

<table>
<tr><th></th><th></th><th colspan="4">Cause for Concern</th><th colspan="3">Neutral</th><th colspan="4">Not Cause for Concern</th></tr>
<tr><th>Item no.</th><th>Item</th><th>5</th><th>4</th><th>3</th><th>2</th><th>1</th><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th><th>0</th></tr>
<tr><td>1</td><td>Overall thought content and clarity</td><td>4</td><td>3</td><td colspan="2">2</td><td colspan="6">1</td><td>0</td></tr>
<tr><td>2</td><td>Concentration</td><td>4</td><td>3</td><td colspan="2">2</td><td colspan="6">1</td><td>0</td></tr>
<tr><td>3</td><td>Feelings of ineffectuality</td><td>3</td><td colspan="4">2</td><td colspan="5">1</td><td>0</td></tr>
<tr><td>4</td><td>Level of need</td><td>4</td><td>3</td><td colspan="3">2</td><td colspan="5">1</td><td>0</td></tr>
<tr><td>5</td><td>Ability to manage symptoms</td><td>4</td><td colspan="2">3</td><td colspan="2">2</td><td colspan="4">1</td><td colspan="2">0</td></tr>
<tr><td>6</td><td>Acceptance of difficulties</td><td>3</td><td colspan="4">2</td><td colspan="5">1</td><td>0</td></tr>
<tr><td>7</td><td>Ability to relax</td><td>2</td><td colspan="4">1</td><td colspan="6">0</td></tr>
<tr><td>8</td><td>Stability of presentation</td><td>4</td><td>3</td><td colspan="4">2</td><td colspan="4">1</td><td>0</td></tr>
<tr><td>9</td><td>Staff Intuition</td><td>3</td><td colspan="5">2</td><td colspan="4">1</td><td>0</td></tr>
<tr><td>10</td><td>Level of functioning</td><td>3</td><td colspan="4">2</td><td colspan="5">1</td><td>0</td></tr>
<tr><td>11</td><td>Energy/get up and go</td><td>3</td><td colspan="4">2</td><td colspan="5">1</td><td>0</td></tr>
<tr><td>12</td><td>Change from normal presentation</td><td>5</td><td>4</td><td>3</td><td colspan="3">2</td><td colspan="4">1</td><td>0</td></tr>
<tr><td>13</td><td>Predictability</td><td>3</td><td colspan="4">2</td><td colspan="5">1</td><td>0</td></tr>
<tr><td>14</td><td>Intensity of symptoms</td><td>5</td><td>4</td><td>3</td><td colspan="2">2</td><td colspan="5">1</td><td>0</td></tr>
<tr><td>15</td><td>Responsibility for self</td><td>3</td><td colspan="4">2</td><td colspan="5">1</td><td>0</td></tr>
<tr><td>16</td><td>Ability to take control</td><td>4</td><td>3</td><td>2</td><td colspan="2">1</td><td colspan="6">0</td></tr>
</table>

Subscale 2 – Adaptive Decision Making

<table>
<tr><th></th><th></th><th colspan="4">Cause for Concern</th><th colspan="3">Neutral</th><th colspan="4">Not Cause for Concern</th></tr>
<tr><th>Item no.</th><th>Item</th><th>5</th><th>4</th><th>3</th><th>2</th><th>1</th><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr>
<tr><td>17</td><td>Thought block</td><td colspan="3">3</td><td colspan="3">2</td><td colspan="4">1</td><td>0</td></tr>
<tr><td>18</td><td>Stream of thought</td><td>4</td><td colspan="2">3</td><td colspan="3">2</td><td colspan="4">1</td><td>0</td></tr>
<tr><td>19</td><td>Flight of ideas</td><td>3</td><td colspan="2">2</td><td colspan="7">1</td><td>0</td></tr>
<tr><td>20</td><td>Poverty of thought</td><td>3</td><td colspan="5">2</td><td colspan="4">1</td><td>0</td></tr>
<tr><td>21</td><td>Understanding of reality</td><td>7</td><td>6</td><td>5</td><td>4</td><td>3</td><td colspan="4">2</td><td>1</td><td>0</td></tr>
<tr><td>22</td><td>Capacity to consent</td><td>3</td><td colspan="4">2</td><td colspan="5">1</td><td>0</td></tr>
<tr><td>23</td><td>Judgement</td><td>8</td><td>7</td><td>6</td><td>5</td><td>4</td><td>3</td><td colspan="3">2</td><td>1</td><td>0</td></tr>
<tr><td>24</td><td>Confusion</td><td>4</td><td colspan="2">3</td><td colspan="3">2</td><td colspan="4">1</td><td>0</td></tr>
<tr><td>25</td><td>Insight</td><td>5</td><td colspan="2">4</td><td colspan="2">3</td><td colspan="3">2</td><td colspan="2">1</td><td>0</td></tr>
<tr><td>26</td><td>Irrational speech</td><td>3</td><td colspan="4">2</td><td colspan="5">1</td><td>0</td></tr>
<tr><td>27</td><td>Overall acceptance of support</td><td>3</td><td colspan="2">2</td><td colspan="7">1</td><td>0</td></tr>
<tr><td>28</td><td>Ability to rationalise</td><td>6</td><td>5</td><td>4</td><td>3</td><td colspan="4">2</td><td colspan="2">1</td><td>0</td></tr>
<tr><td>29</td><td>Speech</td><td>5</td><td>4</td><td colspan="2">3</td><td colspan="3">2</td><td colspan="3">1</td><td>0</td></tr>
<tr><td>30</td><td>Response to hallucinations/ delusions</td><td>4</td><td>3</td><td colspan="2">2</td><td colspan="6">1</td><td>0</td></tr>
</table>

Subscale 3 – Risk of Harm to Self

<table>
<tr><th></th><th></th><th colspan="4">Cause for Concern</th><th colspan="3">Neutral</th><th colspan="4">Not Cause for Concern</th></tr>
<tr><th>Item no.</th><th>Item</th><th>5</th><th>4</th><th>3</th><th>2</th><th>1</th><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr>
<tr><td>31</td><td>Impulsivity</td><td>5</td><td>4</td><td>3</td><td>2</td><td colspan="6">1</td><td>0</td></tr>
<tr><td>32</td><td>Regret of actions during crisis</td><td>6</td><td>5</td><td>4</td><td>3</td><td colspan="2">2</td><td colspan="4">1</td><td>0</td></tr>
<tr><td>33</td><td>Overall risk</td><td>7</td><td>6</td><td>5</td><td colspan="2">4</td><td colspan="2">3</td><td colspan="2">2</td><td>1</td><td>0</td></tr>
<tr><td>34</td><td>Access to lethal means</td><td>6</td><td>5</td><td>4</td><td>3</td><td colspan="2">2</td><td colspan="4">1</td><td>0</td></tr>
<tr><td>35</td><td>Intent to commit suicide</td><td>6</td><td>5</td><td>4</td><td>3</td><td colspan="2">2</td><td colspan="4">1</td><td>0</td></tr>
<tr><td>36</td><td>Regret of suicide attempt</td><td>2</td><td colspan="9">1</td><td>0</td></tr>
<tr><td>37</td><td>Previous attempts at suicide</td><td>4</td><td colspan="2">3</td><td colspan="2">2</td><td colspan="5">1</td><td>0</td></tr>
<tr><td>38</td><td>Risk of suicide</td><td>6</td><td>5</td><td>4</td><td>3</td><td colspan="2">2</td><td colspan="4">1</td><td>0</td></tr>
<tr><td>39</td><td>Risk of harm to self</td><td>5</td><td>4</td><td>3</td><td colspan="2">2</td><td colspan="5">1</td><td>0</td></tr>
<tr><td>40</td><td>Future plans</td><td>5</td><td>4</td><td>3</td><td colspan="2">2</td><td colspan="5">1</td><td>0</td></tr>
</table>

Subscale 4 – Mediating Factors

<table>
<tr><th colspan="2"></th><th colspan="4">Cause for Concern</th><th colspan="3">Neutral</th><th colspan="4">Not Cause for Concern</th></tr>
<tr><th>Item no,</th><th>Item</th><th>5</th><th>4</th><th>3</th><th>2</th><th>1</th><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr>
<tr><td>41</td><td>Social Circumstances</td><td>4</td><td>3</td><td colspan="2">2</td><td colspan="6">1</td><td>0</td></tr>
<tr><td>42</td><td>Protective factors</td><td>4</td><td colspan="2">3</td><td colspan="3">2</td><td colspan="4">1</td><td>0</td></tr>
<tr><td>43</td><td>Resourcefulness</td><td>4</td><td colspan="2">3</td><td colspan="4">2</td><td colspan="3">1</td><td>0</td></tr>
<tr><td>44</td><td>Daily contact with others</td><td colspan="2">3</td><td colspan="2">2</td><td colspan="6">1</td><td>0</td></tr>
<tr><td>45</td><td>Relationships</td><td>3</td><td colspan="4">2</td><td colspan="5">1</td><td>0</td></tr>
<tr><td>46</td><td>Support Networks</td><td colspan="2">3</td><td colspan="3">2</td><td colspan="5">1</td><td>0</td></tr>
</table>

Subscale 5 – Daily Structure

<table>
<tr><th colspan="2"></th><th colspan="4">Cause for Concern</th><th colspan="3">Neutral</th><th colspan="4">Not Cause for Concern</th></tr>
<tr><th>Item no.</th><th>Item</th><th>5</th><th>4</th><th>3</th><th>2</th><th>1</th><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr>
<tr><td>47</td><td>Physical Exercise</td><td>6</td><td>5</td><td>4</td><td>3</td><td colspan="2">2</td><td colspan="4">1</td><td>0</td></tr>
<tr><td>48</td><td>Isolation</td><td>3</td><td colspan="3">2</td><td colspan="6">1</td><td>0</td></tr>
<tr><td>49</td><td>Daily routine</td><td>5</td><td>4</td><td colspan="2">3</td><td colspan="2">2</td><td colspan="4">1</td><td>0</td></tr>
<tr><td>50</td><td>Leisure Activities</td><td>5</td><td>4</td><td colspan="2">3</td><td colspan="2">2</td><td colspan="4">1</td><td>0</td></tr>
<tr><td>51</td><td>Interest/ Enthusiasm</td><td>3</td><td colspan="4">2</td><td colspan="5">1</td><td>0</td></tr>
</table>

Subscale 6 – Risk of Harm to Others

<table>
<tr><th colspan="2"></th><th colspan="4">Cause for Concern</th><th colspan="3">Neutral</th><th colspan="4">Not Cause for Concern</th></tr>
<tr><th>Item no.</th><th>Item</th><th>5</th><th>4</th><th>3</th><th>2</th><th>1</th><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr>
<tr><td>52</td><td>Anger/agitation</td><td>3</td><td colspan="3">2</td><td colspan="6">1</td><td>0</td></tr>
<tr><td>53</td><td>Violence/hostility/ aggression</td><td>4</td><td>3</td><td colspan="3">2</td><td colspan="5">1</td><td>0</td></tr>
<tr><td>54</td><td>Risk of neglect of others</td><td>3</td><td colspan="4">2</td><td colspan="5">1</td><td>0</td></tr>
<tr><td>55</td><td>Family history of suicide</td><td colspan="2">2</td><td colspan="8">1</td><td>0</td></tr>
<tr><td>56</td><td>Risk of harm to others</td><td>4</td><td>3</td><td colspan="2">2</td><td colspan="6">1</td><td>0</td></tr>
<tr><td>57</td><td>Domestic violence</td><td>2</td><td colspan="9">1</td><td>0</td></tr>
</table>

Subscale 7 – Feelings & Affect

<table>
<tr><th></th><th></th><th colspan="4">Cause for Concern</th><th colspan="3">Neutral</th><th colspan="4">Not Cause for Concern</th></tr>
<tr><th>Item no.</th><th>Item</th><th>5</th><th>4</th><th>3</th><th>2</th><th>1</th><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr>
<tr><td>58</td><td>Overall feelings</td><td colspan="2">3</td><td colspan="2">2</td><td colspan="3">1</td><td colspan="4">0</td></tr>
<tr><td>59</td><td>Tearfulness</td><td>3</td><td colspan="4">2</td><td colspan="5">1</td><td>0</td></tr>
<tr><td>60</td><td>Hopelessness</td><td>3</td><td>2</td><td colspan="8">1</td><td>0</td></tr>
<tr><td>61</td><td>Low mood/depression</td><td>4</td><td>3</td><td>2</td><td colspan="3">1</td><td colspan="5">0</td></tr>
</table>

Subscale 8 – Basic Functioning

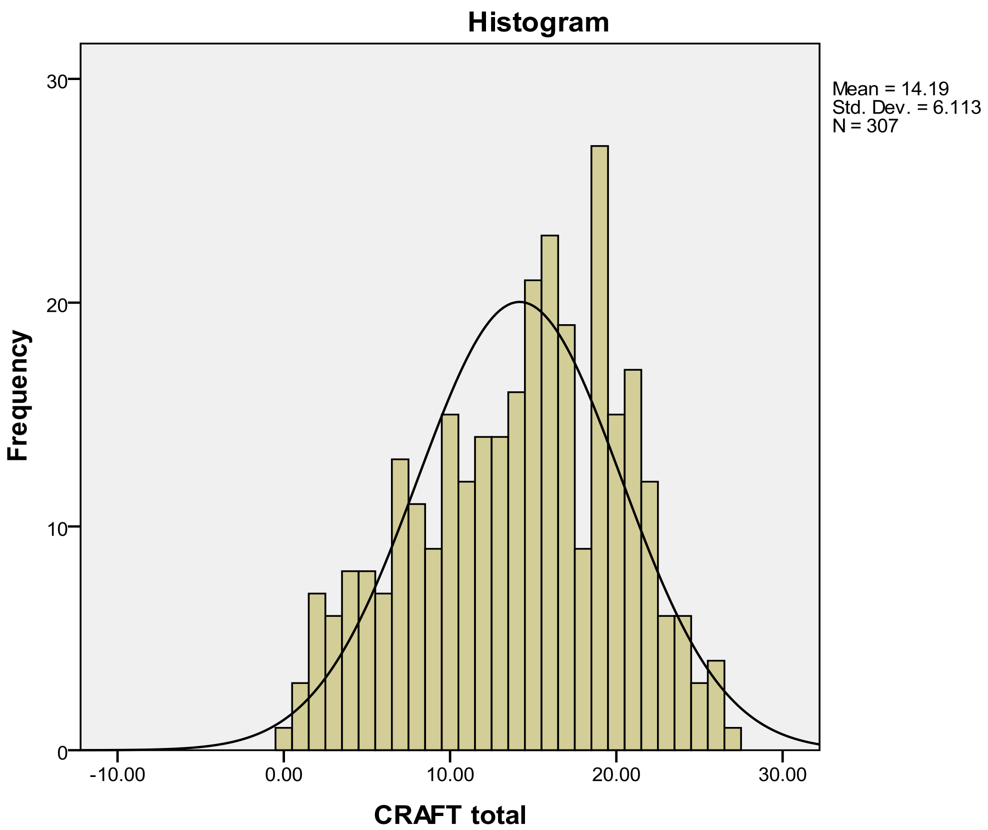
<table>
<tr><th></th><th></th><th colspan="4">Cause for Concern</th><th colspan="3">Neutral</th><th colspan="4">Not Cause for Concern</th></tr>
<tr><th>Item no.</th><th>Item</th><th>5</th><th>4</th><th>3</th><th>2</th><th>1</th><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr>
<tr><td></td><td>Overall appearance</td><td colspan="3">3</td><td colspan="3">2</td><td colspan="4">1</td><td>0</td></tr>
<tr><td></td><td>General wellbeing</td><td colspan="2">3</td><td>2</td><td colspan="3">1</td><td colspan="5">0</td></tr>
<tr><td></td><td>Sleep</td><td>4</td><td>3</td><td colspan="3">2</td><td colspan="5">1</td><td>0</td></tr>
<tr><td></td><td>Appetite</td><td>4</td><td>3</td><td colspan="3">2</td><td colspan="5">1</td><td>0</td></tr>
<tr><td></td><td>Appropriateness of mood</td><td>2</td><td colspan="4">1</td><td colspan="6">0</td></tr>
</table>

Appendix 16 – Descriptive Statistics and Histogram for the 66 Item CRAFT

Descriptive statistics for the CRAFT's total scores

N	Valid	307
	Missing	71
Mean		14.1857
Std. Deviation		6.11342
Range		27.00

Histogram of the CRAFT's total scores



Appendix 17 - Receiver Operator Characteristic Curve Tables

Receiver Operator Curves (ROC) for the different treatment categories as indicated by the Crisis Resolution and Home Treatment teams on the Crisis Measure.

ROC analysis information for the Admission Category

Case Processing Summary

Admission category present		Valid N (list wise)
dimension0	Positive[a]	53
	Negative	243
	Missing	82

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is present.

Coordinates of the Curve

Test Result Variable(s):CRAFT total

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	1.000	.996
1.5000	1.000	.988
2.5000	1.000	.959
3.5000	1.000	.934
4.5000	.981	.905
5.5000	.981	.872
6.5000	.981	.844
7.5000	.981	.790
8.5000	.981	.745
9.5000	.962	.712
10.5000	.925	.663
11.5000	.925	.613
12.5000	.925	.560
13.5000	.887	.510
14.5000	.887	.444
15.5000	.887	.362
16.5000	.849	.280
17.5000	.830	.206
18.5000	.774	.185
19.5000	.585	.119
20.5000	.491	.078
21.5000	.415	.033
22.5000	.264	.016
23.5000	.226	.008
24.5000	.132	.004
25.5000	.075	.004
26.5000	.019	.000
28.0000	.000	.000

ROC analysis information for the Upper Red Category (Visits more than once a day and short stay assessment unit)

Case Processing Summary		
Upper red categories		Valid N (list wise)
dimension0	Positive[a]	23
	Negative	276
	Missing	79

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is present.

Coordinates of the Curve

Test Result Variable(s):CRAFT total

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	1.000	.996
1.5000	1.000	.986
2.5000	1.000	.960
3.5000	1.000	.938
4.5000	1.000	.909
5.5000	1.000	.880
6.5000	1.000	.855
7.5000	1.000	.808
8.5000	1.000	.768
9.5000	.913	.743
10.5000	.870	.696
11.5000	.870	.652
12.5000	.870	.601
13.5000	.870	.551
14.5000	.870	.493
15.5000	.870	.420
16.5000	.826	.344
17.5000	.826	.275
18.5000	.826	.246
19.5000	.652	.163
20.5000	.609	.112
21.5000	.522	.065
22.5000	.304	.040
23.5000	.217	.033
24.5000	.174	.014
25.5000	.087	.011
26.5000	.000	.004
28.0000	.000	.000

ROC analysis information for the Red Category

Case Processing Summary

Red category present		Valid N (listwise)
	Positive[a]	84
	Negative	213
	Missing	81

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is present.

Coordinates of the Curve

Test Result Variable(s):CRAFT total

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	1.000	.995
1.5000	1.000	.986
2.5000	1.000	.953
3.5000	1.000	.925
4.5000	1.000	.887
5.5000	1.000	.850
6.5000	1.000	.817
7.5000	.988	.761
8.5000	.964	.718
9.5000	.905	.700
10.5000	.845	.657
11.5000	.821	.610
12.5000	.786	.563
13.5000	.762	.507
14.5000	.702	.455
15.5000	.607	.399
16.5000	.512	.333
17.5000	.429	.277
18.5000	.417	.244
19.5000	.274	.174
20.5000	.167	.146
21.5000	.071	.113
22.5000	.036	.070
23.5000	.024	.056
24.5000	.012	.033
25.5000	.012	.019
26.5000	.000	.005
28.0000	.000	.000

ROC analysis information for the Amber Category

Case Processing Summary

Amber category present		Valid N (list wise)
dimension0	Positive[a]	77
	Negative	220
	Missing	81

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is present.

Coordinates of the Curve

Test Result Variable(s):CRAFT total

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	1.000	.995
1.5000	1.000	.986
2.5000	.987	.959
3.5000	.948	.945
4.5000	.922	.918
5.5000	.896	.891
6.5000	.870	.868
7.5000	.818	.827
8.5000	.779	.791
9.5000	.779	.750
10.5000	.727	.705
11.5000	.636	.682
12.5000	.571	.645
13.5000	.481	.614
14.5000	.403	.568
15.5000	.325	.505
16.5000	.234	.436
17.5000	.130	.386
18.5000	.104	.359
19.5000	.078	.245
20.5000	.065	.182
21.5000	.026	.127
22.5000	.013	.077
23.5000	.000	.064
24.5000	.000	.036
25.5000	.000	.023
26.5000	.000	.005
28.0000	.000	.000

ROC analysis information for the Green Category

Case Processing Summary

Green category present		Valid N (list wise)
dimension0	Positive[a]	81
	Negative	212
	Missing	85

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is present.

Coordinates of the Curve

Test Result Variable(s):CRAFT total

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	.988	1.000
1.5000	.963	1.000
2.5000	.889	.995
3.5000	.852	.981
4.5000	.790	.967
5.5000	.716	.958
6.5000	.654	.948
7.5000	.556	.925
8.5000	.481	.901
9.5000	.444	.873
10.5000	.407	.821
11.5000	.370	.778
12.5000	.309	.741
13.5000	.284	.684
14.5000	.235	.627
15.5000	.160	.561
16.5000	.086	.486
17.5000	.049	.410
18.5000	.025	.382
19.5000	.000	.269
20.5000	.000	.198
21.5000	.000	.127
22.5000	.000	.071
23.5000	.000	.052
24.5000	.000	.028
25.5000	.000	.019
26.5000	.000	.005
28.0000	.000	.000

ROC analysis information for the Discharge Category

Case Processing Summary

Discharge		Valid N (list wise)
dimension0	Positive[a]	56
	Negative	241
	Missing	81

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is Present.

Coordinates of the Curve

Test Result Variable(s):CRAFT total

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	1.000	.996
1.5000	.964	.996
2.5000	.911	.979
3.5000	.857	.967
4.5000	.821	.942
5.5000	.732	.929
6.5000	.643	.921
7.5000	.518	.896
8.5000	.411	.876
9.5000	.357	.851
10.5000	.339	.797
11.5000	.339	.747
12.5000	.304	.701
13.5000	.268	.651
14.5000	.214	.598
15.5000	.143	.531
16.5000	.089	.452
17.5000	.054	.382
18.5000	.018	.357
19.5000	.000	.249
20.5000	.000	.187
21.5000	.000	.124
22.5000	.000	.075
23.5000	.000	.058
24.5000	.000	.033
25.5000	.000	.021
26.5000	.000	.004
28.0000	.000	.000

Appendix 18 - CRAFT

Crisis Risk and Adaptive Functioning Tool

Assessment of adaptive community functioning in the context of acute mental health crisis

The CRAFT is for use by mental health professionals assessing patients for acute mental health crisis in the community setting. The CRAFT is only to be completed by those trained in the how to use the tool with the flexible rating system. To accompany the use and scoring of the CRAFT is the CRAFT manual which contains a full descriptive definition for each of the items.

Initial Checklist Items (please circle) If scored 'yes', please consider whether these risks can be safely managed in the community setting.			
Risk of arson	Yes No	Risk of homicide	Yes No

Component 1 – Crisis Recovery Indicators												
No.	Item	Cause for concern				Balanced concern			Not cause for concern			
1	**Overall thought content and clarity**	5	4	3	2	1	0	1	2	3	4	5
2	**Concentration**	5	4	3	2	1	0	1	2	3	4	5
3	**Feelings of ineffectuality**	5	4	3	2	1	0	1	2	3	4	5
4	**Level of need**	5	4	3	2	1	0	1	2	3	4	5
5	**Ability to manage symptoms**	5	4	3	2	1	0	1	2	3	4	5
6	**Acceptance of difficulties**	5	4	3	2	1	0	1	2	3	4	5
7	**Ability to relax**	5	4	3	2	1	0	1	2	3	4	5
8	**Stability of presentation**	5	4	3	2	1	0	1	2	3	4	5
9	**Staff Intuition**	5	4	3	2	1	0	1	2	3	4	5
10	**Level of functioning**	5	4	3	2	1	0	1	2	3	4	5
11	**Energy/get up and go**	5	4	3	2	1	0	1	2	3	4	5
12	**Change from normal presentation**	5	4	3	2	1	0	1	2	3	4	5
13	**Predictability**	5	4	3	2	1	0	1	2	3	4	5
14	**Intensity of symptoms**	5	4	3	2	1	0	1	2	3	4	5
15	**Responsibility for self**	5	4	3	2	1	0	1	2	3	4	5
16	**Ability to take control**	5	4	3	2	1	0	1	2	3	4	5
Total Transformed Score **(use transformation table)**		/98										

Component 1 – Total Transformed Score Guideline (please tick)	
Low	**0-29**
Medium	**30-57**
High	**58-98**

Component 2 – Adaptive Decision Making												
No.	Item	Cause for concern				Balanced concern			Not cause for concern			
17	**Thought block**	5	4	3	2	1	0	1	2	3	4	5
18	**Stream of thought**	5	4	3	2	1	0	1	2	3	4	5
19	**Flight of ideas**	5	4	3	2	1	0	1	2	3	4	5
20	**Poverty of thought**	5	4	3	2	1	0	1	2	3	4	5
21	**Understanding of reality**	5	4	3	2	1	0	1	2	3	4	5
22	**Capacity to consent**	5	4	3	2	1	0	1	2	3	4	5
23	**Judgement**	5	4	3	2	1	0	1	2	3	4	5
24	**Confusion**	5	4	3	2	1	0	1	2	3	4	5
25	**Insight**	5	4	3	2	1	0	1	2	3	4	5
26	**Irrational speech**	5	4	3	2	1	0	1	2	3	4	5
27	**Overall acceptance of support**	5	4	3	2	1	0	1	2	3	4	5
28	**Ability to rationalise**	5	4	3	2	1	0	1	2	3	4	5
29	**Speech**	5	4	3	2	1	0	1	2	3	4	5
30	**Response to Hallucinations/Delusions**	5	4	3	2	1	0	1	2	3	4	5
Total Transformed Score **(use transformation table)**		/77										

Component 2 – Total Transformed Score Guideline(please tick)	
Low	**0-18**
Medium	**19-35**
High	**36-77**

Component 3 – Risk of harm to self												
No.	Item	Cause for concern				Balanced concern			Not cause for concern			
31	**Impulsivity**	5	4	3	2	1	0	1	2	3	4	5
32	**Regret of actions during crisis**	5	4	3	2	1	0	1	2	3	4	5
33	**Overall risk**	5	4	3	2	1	0	1	2	3	4	5
34	**Access to lethal means**	5	4	3	2	1	0	1	2	3	4	5
35	**Intent to commit suicide**	5	4	3	2	1	0	1	2	3	4	5
36	**Regret of suicide attempt**	5	4	3	2	1	0	1	2	3	4	5
37	**Previous attempts at suicide**	5	4	3	2	1	0	1	2	3	4	5
38	**Risk of suicide**	5	4	3	2	1	0	1	2	3	4	5
39	**Risk of harm to self**	5	4	3	2	1	0	1	2	3	4	5
40	**Future plans**	5	4	3	2	1	0	1	2	3	4	5
Total Transformed Score (use transformation table)		/58										

Component 3 – Total Transformed Score Guideline (please tick)	
Low	**0-18**
Medium	**19-35**
High	**36-58**

Component 4 – Protective/Mediating Factors												
No.	Item	Cause for concern				Balanced concern			Not cause for concern			
41	**Social Circumstances**	5	4	3	2	1	0	1	2	3	4	5
42	**Protective factors**	5	4	3	2	1	0	1	2	3	4	5
43	**Resourcefulness**	5	4	3	2	1	0	1	2	3	4	5
44	**Daily contact with others**	5	4	3	2	1	0	1	2	3	4	5
45	**Relationships**	5	4	3	2	1	0	1	2	3	4	5
46	**Support Networks**	5	4	3	2	1	0	1	2	3	4	5
Total Transformed Score (use transformation table)		/33										

Component 4 – Total Transformed Score Guideline (please tick)	
Low	**0-10**
Medium	**11-19**
High	**19-33**

Component 5 – Daily Structure												
No.	Item	Cause for concern				Balanced concern			Not cause for concern			
47	**Physical Exercise**	5	4	3	2	1	0	1	2	3	4	5
48	**Isolation**	5	4	3	2	1	0	1	2	3	4	5
49	**Daily routine**	5	4	3	2	1	0	1	2	3	4	5
50	**Leisure Activities**	5	4	3	2	1	0	1	2	3	4	5
51	**Interest/Enthusiasm**	5	4	3	2	1	0	1	2	3	4	5
Total Transformed Score (use transformation table)		/29										

Component 5 – Total Transformed Score Guideline(please tick)	
Low	**0-9**
Medium	**10-19**
High	**20-29**

Component 6 – Risk of harm to others												
No.	Item	Cause for concern				Balanced concern			Not cause for concern			
52	**Anger/agitation**	5	4	3	2	1	0	1	2	3	4	5
53	**Violence/hostility/aggression**	5	4	3	2	1	0	1	2	3	4	5
54	**Risk of neglect of others**	5	4	3	2	1	0	1	2	3	4	5
55	**Family history of suicide**	5	4	3	2	1	0	1	2	3	4	5
56	**Risk of harm to others**	5	4	3	2	1	0	1	2	3	4	5
57	**Domestic violence**	5	4	3	2	1	0	1	2	3	4	5
Total Transformed Score (use transformation table)		/24										

Component 6 – Total Transformed Score Guideline (please tick)	
Low	**0-5**
Medium	**6-11**
High	**12-24**

Component 7 – Feelings/Affect												
No.	Item	Cause for concern				Balanced concern			Not cause for concern			
58	**Overall feelings**	5	4	3	2	1	0	1	2	3	4	5
59	**Tearfulness**	5	4	3	2	1	0	1	2	3	4	5
60	**Hopelessness**	5	4	3	2	1	0	1	2	3	4	5
61	**Low mood/depression**	5	4	3	2	1	0	1	2	3	4	5
Total Transformed Score **(use transformation table)**		/30										

Component 7 – Total Transformed Score Guideline (please tick)	
Low	**0-10**
Medium	**10-19**
High	**20-30**

Component 8 – Basic Functioning												
No.	Item	Cause for concern				Balanced concern			Not cause for concern			
62	**Overall appearance**	5	4	3	2	1	0	1	2	3	4	5
63	**General wellbeing**	5	4	3	2	1	0	1	2	3	4	5
64	**Sleep**	5	4	3	2	1	0	1	2	3	4	5
65	**Appetite**	5	4	3	2	1	0	1	2	3	4	5
66	**Appropriateness of mood**	5	4	3	2	1	0	1	2	3	4	5
Total Transformed Score **(use transformation table)**		/28										

Component 8 – Total Transformed Score Guideline (please tick)	
Low	**0-8**
Medium	**9-16**
High	**17-28**

Scale Summary Sheet

Service User's Name:______________________MPI no.:__________________

Rater's Name:________________________

High								
Medium								
Low								
	Component **1** Crisis Recovery Indicators	Component **2** Adaptive Decision Making	Component **3** Risk of Harm to Self	Component **4** Protective/ Mediating Factors	Component **5** Daily Structure	Component **6** Risk of Harm to Others	Component **7** Feelings/ Affect	Component **8** Basic Functioning

Assessment Outcome:____

Treatment Focus:____

Appendix 19 – Number of measures completed by staff participants