Development and testing of a measure of relational depth in counselling and psychotherapy

By

Sue Price

A Doctoral Thesis
Completed in order to meet the criteria for the title of Doctor of Philosophy of the University of Strathclyde.

Sue Price

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: (Sue Price)*

Date:

*Please note that I, the author, was formerly known as Sue Wiggins at the commencement of the PhD project presented here and I changed my surname, by deed poll, to my birth name of Price before the final submission of this PhD project.

Table of Contents

Abstract

Mearns and Cooper (2005), describe relational depth as 'a state of profound contact and engagement between two people…' (p. xii). Regarding the relational aspects of therapy, the wealth of research literature suggests that there is a common set of factors to do with the therapeutic relationship across all approaches of therapy which leads to therapeutic improvement.

At the time of writing, there is only one measure designed to assess relational depth; the Relational Depth Inventory (RDI). However, the RDI has not been as thoroughly tested for validity and reliability as have other relational measures. No quantitative studies have yet been conducted that explore the association between relational depth and therapeutic outcome that use a relational depth inventory and traditional outcome measures in a practice base. Therefore, the main objective of this research project was to explore, improve and evidence the reliability and validity of the RDI.

Using traditional statistics, the first study found evidence that the RDI is essentially unidimensional but composed of a number of elements especially those associated with deep relational experiences such as love, connectedness and respect. Gender was seen to have a limited roll as a determinant of relational depth events. In the second study, Rasch analyses evidenced that the revised RDI, the RDI-R, had excellent reliability. The RDI-R was found to be largely undimensional but that a significant minority of persons experiencing high strength of relational depth were not targeted. In the third study, presence of relational depth, as assessed by the RDI-R, during helpful moments in therapy, was found to be significantly associated with therapeutic improvement.

The results imply that the RDI-R has a range of different kinds of validity evidence, although there is room for improvement and that relational depth is related to therapeutic improvement.

Chapter 1. Introduction to this Research Project

1.1 Introduction to Chapter

This, the first chapter in this document, introduces the reader to this whole research project. It does so by first outlining why the research presented here might be important in terms of implications to therapy practice and improvement. It also gives personal reasons as to why I embarked on such a research project. I then go on to document previous research conducted by myself as part of an MSc project which formed the basis for the whole of this PhD research. This part is designed to clarify exactly where the MSc finished and where the PhD began. I outline the main objectives for carrying out this research as well as setting out any expectations of what this research will result in. I introduce the reader to myself, the researcher, and then finally define terms used throughout the project. I also include a part on humanistic values and the scientific method which explains to the reader why these two aspects are often, apparently, seen to be at conflict with each but why they also both play a part in this research project.

1.2 Why this research project might be important

1.2.1 Personal Reasons

When I first read the book Working at Relational Depth in Counselling and Psychotherapy (Mearns and Cooper, 2005), I knew I wanted to find a way to be able to tap into those particular encounters which the authors describe as ‘moments’, ‘times’ or ‘experiences’ of relational depth. More specifically I wanted to be able to create an inventory (a measure or questionnaire) designed to assess such moments, times and experiences. Having studied psychology for my first degree, I had become fascinated by how inventories and measures were created. Despite my enthusiasm I

knew that creating an inventory designed to assess relational depth would be a huge task because it required attempting to quantify a deep, meaningful and uniquely human experience, an experience that up until then had only been researched qualitatively. This posed the question: How does one quantify such a deep and personal an experience as relational depth? In other words, how does one create a measure designed to assess moments of relational depth in therapy? Such questions exited me and thus gave me the motivation to set about finding ways in which I might create such a measure.

1.2.2 Implications for therapy practice and outcome.

Another valid reason for conducting this research concerns the implications it may have for therapy. An inventory designed to assess relational depth may prove a useful tool in the research of the therapy professions. Firstly, such an inventory can be used to assess the relation between relational depth and outcome and this in turn may help inform practice as to how the therapist may engage in self-development. For example, if evidence is found that relational depth is associated with outcome, this may imply that therapists need to be open to the possibility that clients may want and need such deep engagement to be able to heal, change, improve and make full use of therapy. Secondly, a relational depth inventory may also be used in training environments where students in client role are able to rate their peers taking on therapist role. This may help in creating an awareness, especially in Person-centred students, of the experiences of person-centred therapy.

1.3 Previous research by current author.

The development of the RDI inventory began outside of this PhD project and was part of an MSc completed in July 2007. This PhD project officially began only three months later at the beginning of the academic year in 2007. To enable a better understanding of this PhD project it is important that the steps used in the aforementioned MSc project are documented in order that it is clear what validation steps were already taken before the beginning of the PhD research. It is also important that clarification is made as to what extent the MSc project contributed towards the development of the RDI and where the PhD resumed such development. In light of this, this subsection briefly documents which parts of the RDI development were part of my MSc study and which parts of it were resumed by the current PhD project.

The first steps in creating the Relational Depth Inventory (RDI) were carried out as part of an MSc project (Wiggins, 2007) which largely consisted of creating questionnaire items from a collection of qualitative data. The qualitative data consisted of counsellor and client descriptions of relational depth experiences taken from two sources. The first source was archival data in the form of a list of descriptions and phrases from brainstorming exercises in which counsellors and trainee counsellors, describe their experience of moments of relational depth following a lecture given on that subject. In total, 261 descriptions of relational depth were collected during these exercises. These descriptions were utilised in the current study and, as such, formed a large part of its raw data.

The second source of raw data was a partially analysed data set from research conducted by Knox (2007). This data was the result of interviews where participants were 14 clients of Person-Centred counselling. All participants were counsellors or trainee counsellors and consisted of 9 women and 5 men. Knox (2007) stated that participants were from a range of ages, ethnicity and backgrounds and were from different parts of the UK. Participants were selected by advertising, talking to groups and from the researchers own network. In each of the 14 interviews participants described their experiences of specific moments of relational depth and this resulted in over 100 descriptions.

All descriptions were subjected to grounded theory analysis (Rennie, Phillips, & Quartaro, 1988) and resultant categories guided questionnaire items (see Table 1.1). Items were created like this in order to contribute content-related validity evidence. It also served as a form of theory generation or theory addition in that it used many different opinions as to what relational depth consisted of. This therefore provided what might be termed as content-evidence for validity. Developing the RDI in this way provided a more objective way of creating inventory items that would not otherwise be achieved if just one researcher created items from their own limited view of relational depth. This preliminary stage of the validation process could be seen as what Rowe and Mahar (2006) call the 'definitional stage'. This is mainly because it involved theory and evidence to provide descriptions (from definitions) of the construct as well as having a more 'test-centered' focus.

Table 1.1 Grounded theory analysis of clients' and counsellor descriptions of relational depth showing domains and categories.

Domain Level	Level 1 Categories	Level 2 Subcategories	Level 3 Subcategories	N*
Experience of Relationship				83
" "	**Connectedness**			43
" "	"	Connectedness		16
" "	"	"	Meeting	7
" "	"	Spiritual Connectedness		5
" "	"	Love		10
" "	"	Intimacy		12
" "	"	"	Intimacy in the Moment[a]	3
" "				
" "	**Mutuality**			26
" "	"	Mutuality		6
" "	"	Reciprocity		5
" "	"	Attunement		8
" "	"	Commonality		4
" "	"	Equality		3
" "				
" "	**Security**			14
" "	"	Safety		4
" "	"	"	Safety in the Moment[a]	1
" "	"	Privacy		5
" "	"	"	Privacy in the Moment[a]	1
" "	"	Trust		5
Experience of Self				173
" "	**Heightened Self**			56
" "	" "	Expansiveness		34
" "	" "	"	Expansiveness	3
" "	" "	"	Feeling free	9
" "	" "	"	Feeling spontaneous	6
" "	" "	"	Openness	9
" "	" "	"	Flow	7
" "	" "			
" "	" "	Feeling Spiritual		19
" "	" "	" "	Spiritual	6
" "	" "	" "	Altered state	4
" "	" "	" "	Transcendent	4
" "	" "	" "	Magical	2
" "	" "	" "	Enlightened	3
" "	" "			
" "	" "	Feeling happy		3
" "				

"	"	**Invigorated Self**				46
"	"	"	"	Invigorated		22
"	"	"	"	"	Energising	11
"	"	"	"	"	Empowering	2
"	"	"	"	"	Vitalising	7
"	"	"	"	"	Exhilarating	2
"	"	"	"			
"	"	"	"	Risky/afraid		11
"	"	"	"	Intense		13
"	"	"	"	"	Intense feelings	3
"	"	"	"	"	Excited	8
"	"	"	"	"	Angry	1
"	"	"	"	"	Sexual	1
"	"					
"	"					
"	"	**Immersed self**				25
"	"	"	"	Aware of Experience		4
"	"	"	"	Immersed in Feelings		21
"	"	"	"	"	"	Immersion 9
"	"	"	"	"	"	Feelings 3
"	"	"	"	"	"	Deep feelings 4
"	"	"	"	"	"	Opposing feelings 3
"	"	"	"	"	"	Weird feelings 2
"	"					
"	"					
"	"	**True self**				46
"	"	"	"	Authenticity		23
"	"	"	"	"	Being real	9
"	"	"	"	"	Vulnerable	6
"	"	"	"	"	True self in the moment	8
"	"	"	"	In touch with self		11
"	"	"	"	Self-valuing		12
Experience of towards other						62
"	"	"	**UPR**[b]			16
"	"	"	"	Respecting other		9
"	"	"	"	Other respecting me[a]		3
"	"	"	"	Prizing		4
"	"	"				
"	"	"				
"	"	"	**Trust**			11
"	"	"	"	Trust of other		10
"	"	"	"	Other trusting me[a]		1
"	"	"				
"	"	"	**Being Available**			17
"	"	"	"	"	Being available to other	10
"	"	"	"	"	Other available to me[a]	7
"	"	"				

"	"	"	**Empathy**			9
"	"	"	"		Empathy for other	5
"	"	"	"		Other being empathic[a]	4
"	"	"				
"	"	"	**Other being real[a]**			9
"	"	"	"	"	"	Being real[a]	7
"	"	"	"	"	"	Being real in the moment[a]	2
Experience of Atmosphere						49
"	"		**Dynamism**			12
"	"		"		Dynamism	8
"	"		"		Dynamism in the moment	4
"	"					
"	"		**Peace**			22
"	"		"		Timeless	12
"	"		"		Stillness	10
"	"					
"	"					
"	"		**Significance**			15
"	"		"		Meaningful	5
"	"		"		Awesome	4
"	"		"		Inexplicable	3
"	"		"		Unique	3

N* Number of descriptions contributing to each domain, category or subcategory. [a]created due to addition of Data Set 2 [b] Unconditional Positive Regard

After an initial set of 64 items had been created (from descriptions as mentioned above), the initial RDI was designed in a way so that there would be two types of observation that would later be utilized for content-related evidence. Firstly, descriptions of important events were collected from participants. Secondly, the 64 relational depth items were each presented with a Likert scale for participants to rate their important event. The Working Alliance Inventory was added to the questionnaire so that the similar construct of working alliance could be investigated as to what extent it correlated with the initial relational depth inventory. In the MSc study 80 participants had responded to the 64-item RDI and some initial analyses were conducted. This is where the MSc ended. To obtain more statistical power for reliability and validity studies, more participants were needed. The PhD study began

by obtaining more participants. For the PhD study, a further 263 participants completed the 64-item RDI making a total of 343. This first study of the PhD is documented in Study 1.

1.4 Main research objectives

1.4.1 Improve Reliability and Validity of the Relational Depth Inventory (RDI)

One of the main aims of this research project were to refine the RDI by testing and improving the reliability and validity of the 64-Item Relational Depth Inventory (RDI) which was designed to assess clients' experiences of relational depth during a specific event. The specific aims regarding reliability and validity are set out below.

1. To test the pre-existing Relational Depth Inventory (RDI) for validity evidence. This will include providing face-value related evidence, content-related evidence and criterion-related evidence.
2. To investigate the characteristics of the construct of relational depth, including any underlying dimension/s.
3. To explore if the experience of relational depth, as reported by clients and therapists, was consistent with research already carried out.
4. To investigate whether there were any variables that affected the likelihood of relational depth being present. It was intended that gender and age would be explored.

1.4.2 Explore the relation between relational depth and improvement

The second main objective of this project was to explore the relationship between relational depth and therapeutic improvement. Although arguably this had already

been done before using qualitative methodology (Knox, 2008) and had been done using quantitative methods (Leung, 2009) the relation between improvement and relational depth had never been explored quantitatively whilst using tried and tested outcome measures such as the Clinical Outcome Routine Evaluation (CORE) inventory. Therefore this part of the project set out to do the following

1. Investigate the association between relational depth and outcome/improvement and to do so as objectively as possible whilst also being ecologically valid. This would mean using quantitative methodology but, at the same time, having participants to take part in this research in an environment that is natural to them rather than in the clinical surroundings of a laboratory. This part of the research would be therefore practice-based which means having no control group. This would be done using a shorter version of the RDI which was intended to be developed in Study 1.

1.4.3 Explore the relation between relational depth and working alliance

The third objective of this research is to explore the relation between relational depth and working alliance. This would help establish if there were any considerable differences between the two constructs. The main reason for this is that both these constructs attempt to explain aspects of the relationship between practitioner and client. Whilst there are relational constructs that are very similar to relational depth , more similar than the working alliance construct, none have been researched and documented quite as much as working alliance. Furthermore, there is in existence a 36-item Working Alliance Inventory (WAI) which has been developed by Adam

Horvath and Leslie Greenberg (e.g. Horvath & Greenberg, 1986). This has later been developed by Tracey and Kokotovic (1989) to create a short 12-item version of the WAI (WAI-S) and more recently Hatcher & Gillaspy (2006) created a revised short form of the WAI (WAI-SR). The WAI has arguably been more tried and tested than most other relational measures and this is the reason for using it in the current research project (see Chapter 2 for a review of WAI research and Chapter 8 for a complete description of the particular version of the WAI used in this project).

1.5 The Researcher

I am a BACP Accredited therapist and the way I practice draws heavily from the works of Rogers (1951, 1957). Consequently, I consider myself to be Person-centred in my approach as a therapist. As well as this, I believe I have encountered moments of relational depth with many clients. However, these times have been rare and never seem to have come easily. I am aware there are some (e.g. Wilders, 2007) who see the theory behind relational depth as being in conflict with the Person-centred approach. My own experience and belief is that the theory underpinning relational depth encompasses and embraces the conditions of Rogers (1957) that he claimed were necessary for therapy.

As already mentioned, my approach to therapy is Person-centred. However, my own Person-centred approach is not limited to therapy. In every step of the research conducted here, I consciously always had in mind my humanistic convictions and undertook this project with a person-centred attitude. Carrying out research with a Person-centred approach, in my experience, means to not direct the research with a heavy hand but rather to be guided by the information that I already had in my

possession. The particular information I mean is the collection of descriptions of relational depth experiences and the resulting 64-item Relational Depth Inventory. In addition, it felt important to me that if there was any strong direction in this research, it should only be in regard to particular research techniques. For example, when developing any type of inventory, tried and tested quantitative techniques are often carried out (such as validity and reliability techniques) and this is the case with the research in this project. In carrying out this research in a more person-centred way I believe I have not so much directed the research but let the research direct itself to a considerable degree; I therefore see myself more as a facilitator to the whole research project.

1.6 Apparent incongruities: Humanistic values and the scientific method

As this research project draws heavily on quantitative methodology, it may be regarded by some as not being humanistic. A student of Ambika Connelly said 'I'm quite happy for it [relational depth] to be elusive…. There's a fear of, you know – it's like kind of butterfly catching, isn't it? – there's a fear of catching something very beautiful and trying to define what it is. And then, and in that process, losing what it is' (Connelly, 2009).

There is the argument that using scientific methods, as in measuring, quantifying and empirically examining human experience is not humanistic simply because it is seen to advocate reductionism and objectification. Paradoxically, I consider myself to be humanistic and I also consider myself to be an advocate of the scientific method. However, I fully acknowledge the limitations of using the scientific method when

applied to the human experience. The quote below by Charles Spearman (Spearman, 1904) puts it succinctly

> All knowledge – beyond that of bare isolated occurrence – deals with uniformities. Of the latter, some few have a claim to be considered absolute, such as mathematical implications and mechanical laws. But the vast majority are only partial; medicine does not teach that smallpox is inevitably escaped by vaccination, but that it is so generally; biology has not shown that all animals require organic food, but that nearly all do so; in daily life, a dark sky is no proof that it will rain, but merely a warning; even in morality, the sole categorical imperative alleged by Kant was the sinfulness of telling a lie, and few thinkers since have admitted so much as this to be valid universally. In psychology, more perhaps than in any other science, it is hard to find absolutely inflexible coincidences; occasionally, indeed, there appear uniformities sufficiently regular to be practically treated as laws, but infinitely the greater part of the observations hitherto recorded concern only more or less pronounced *tendencies* of one event or attribute to accompany another. (p.72)

Spearman (1904) was simply acknowledging the limitations of using the scientific method when attempting to measure human experience.

Similarly, Carl Rogers (1951, when introducing one of his most noteworthy publications, *Client-Centered Therapy,* said that the book was about his colleagues and himself as they undertake scientific analysis. He also stated it was about conflicts in that he and his colleagues felt strongly that therapeutic process was rich

in ‘shading, complexities and subtleties’ (p. xi) and that the ‘scientific finding…is cold, lifeless, and lacking in the fullness of the experience.’ (p. xi). However, Rogers (1951) further added that the book also expressed his and fellow colleagues’ conviction

> …that though science can never make therapists, it can help therapy; that though the scientific finding is cold and abstract, it may assist us in releasing forces that are warm, personal, and complex; and that though science is slow and fumbling, it represents the best road we know to the truth, even in so delicately intricate an area as that of human relationships. (p. xi)

As a humanistic therapist, I believe that one cannot fully quantify the human experience and I hold the viewpoints of the psychologists quoted above; the scientific method, or any method for that matter, is cumbersome, fallible and imperfect. I firmly believe that no single method of research is superior to another. However, I do believe that certain methods are better than others for certain jobs. Using a screwdriver to hammer a nail in is not nearly as effective as using a hammer. Similarly, in conducting the current research, using qualitative methodology would not have carried out the job in hand nearly as effectively as a quantitative methodology. Therefore, I believe I have used the best tool for the job in hand.

1.7 Expectations of this PhD Project

The research presented here for this whole PhD project is largely exploratory. By this I mean that I do hold any particular predictions. However, whilst I neither

predict nor hold any particular hypotheses for the results of this research, I do have some expectations of what the research may find.

For example, I do expect that in any factor analyses of questionnaire items there will be evidence that empathy, unconditional positive regard and genuineness are the main ingredients of relational depth. With regard to any Rasch analyses which look at item difficulty I expect that the items which assess client empathy and spirituality will be the hardest items to endorse because these experiences appear to be on a higher level of relational depth than any other. Evidence of this comes from Knox (2008) who in a qualitative study found that clients reported that the experience of relational depth regarding the relationship (as opposed to self or therapist) was on 'a different dimension/mystical/spiritual' (p.186). With regard to relational depth and client improvement, I expect that significant moments in therapy which have a presence of relational depth will be related to improvement. However I do not expect this to be large correlation given that clients are only being asked to describe one significant event. In other words I do not expect one significant event to be related to client improvement simply because it is only one event. I anticipate that it will be the ongoing relationship or working alliance between client and therapist which will predict client improvement more than a significant event.

1.8 Defining terms

1.8.1 Counselling and Psychotherapy

In this thesis I use the terms 'counselling', 'psychotherapy' and 'therapy' to mean the same thing and therefore I use these terms interchangeably. I see therapy, at least within this project, to be a confidential relationship where the client is free to

disclose to their therapist things that they may not find able to do with anybody else. In addition, my view of therapy is that it is a process where a client meets with a therapist with the express intention of improving their psychological health by having a better understanding of themselves.

1.8.2 Psychological Testing

Kaplan and Saccuzzo (2009) state that psychological testing 'refers to all the possible uses, applications, and underlying concepts of psychological and educational tests.' (p.9). A psychological test, therefore, attempts to measure or assess a psychological construct. Psychological constructs are typically depression, anxiety, intelligence and in the case of this project, relational depth. In this vein I include the assessment of relational depth as a form of psychological testing. The term *testing* is perhaps an unfortunate one especially since I aim to approach this project with a Person-centred attitude. In other words, although any Relational Depth Inventory I create will be designed to test for presence of relational depth, I do not hold any greater or lesser value to higher or lower presence of relational depth.

The term *test* is used in this project to encompass such terms as scale, measure, instrument and inventory. I also use all of these terms interchangeably. In doing so I also refer to the 'testing' of participants. Similar to above I use the term simply because it is a recognized term and therefore using such a term will aid clarity.

1.8.3 Relational Depth

I use the most recent and comprehensive definition of relational depth as a basis for this study. In their publication, Mearns and Cooper (2005) discuss at length

what they mean by their definition. They propose there are two aspects to relational depth, which are related. They use the term relational depth to describe specific moments of encounter as well as a particular quality of a relationship. With regard to specific moments of encounter, they use the words '*moments*' '*times*' and '*experiences*' to describe these particular, relatively short-lived experiences. With regard to the particular quality of the relationship, they describe this as having more of an enduring quality where there is 'an enduring sense of contact and interconnection between two people' (p.xii). Although slightly different, these two aspects of relational depth still have at their core the emphasis on relationship. To aid clarity and reduce confusion, I have focused on one definition. Consequently, for the purposes of this research I will use the definition of relational depth that has been put forward by Mearns and Cooper (2005), which is as follows:

> *A state of profound contact and engagement between two people, in which each person is fully real with the Other, and able to understand and value the Other's experiences at a high level' (p.xii).*

Despite there being the two aspects to relational depth described above, the current project focuses on one aspect which sees relational depth as a moment, time or experience of relational depth. Consequently, the relational depth measure which is being developed is designed to assess the presence of relational depth during a particular moment in therapy.

1.8.4 Terminology of Participants' Endorsing of Items.

A large portion of this project focuses on Rasch analysis, a form of Item Response Theory (IRT) although very different from most IRT approaches. Rasch analysis methods were originally developed for educational purposes in order to improve the reliability and validity of tests designed to assess students' ability across a number of different subjects (e.g. maths) (Bond and Fox, 2007). Because of this, and due to the particular analyses, terms such as 'item difficulty' and more importantly, 'person ability' are commonly used. However, in the current study, person ability is inappropriate when referring to participants' responses on the Relational Depth Inventory. This is because a measure of relational depth presence is not assessing an ability of any kind. Furthermore, in the spirit of the person-centred approach, I wish to not be judgmental or disrespectful in any way. The problem is then how to conveniently phrase the equivalent to person ability without using the word *ability*. The term *person agreeableness* seems to be the most appropriate phrase although even this is not perfect because the item anchor points are not on an 'agree/disagree scale.' However, it is convenient and non-disrespectful. I only use the term person agreeableness when referring to the inventory being developed here. When reviewing various Rasch literature, I will refer to person ability where appropriate especially if the studies being cited use the same terms. In any case, person ability and person agreeableness can be thought of as a person's capacity to endorse an item which in turn is representative of the amount of trait being measured that a test-taker possesses.

1.9 Chapter Summary

This chapter has introduced the reader to the current research project by outlining why this research might be important and has documented the main research objectives and expectations. It is hope that this has now prepared the reader for what is to follow. The next three chapters reviews the literature on relational processes, relevant epistemology and relevant methodological issues. After these literature reviews chapters are given which outline in detail the actual research conducted as part of this project.

Chapter 2: A Review of Relational Depth and Relational Processes Literature

2.1 Introduction to Chapter

This chapter, Chapter 2, documents in detail all the relevant literature concerning *relational depth*. It describes in detail the theory underlying relational depth and looks at how the term came into being before focusing on other relational constructs which are comparable to the concept relational depth. It then goes on to document measures of the client-therapist relationship. This is in order that the Relational Depth Inventory (RDI) can be put into context and compared with pre-existing measures. The chapter also includes literature concerning the relation between relational depth and outcome/improvement as well as other relational aspects and outcome (e.g. working alliance and outcome). The purpose of documenting this literature is so that the concept of relational depth and the RDI can be put into context with other relational aspects and other relational measures.

2.2 What is Relational Depth

2.2.1 Development of the term 'relational depth'.

Mearns (1996) first coined the term 'relational depth' when discussing Rogers' (1957) core conditions (refer to p. 27 for the list of Rogers' core conditions). More specifically, Mearns (1996) talked about the first of the core conditions, psychological contact. He suggested that psychological contact is not an all or nothing matter but rather on a spectrum and stated that 'the interaction between counsellor and client will move around the contact spectrum, at times engaging very deeply and on other occasions much more superficially' (Mearns, 1997, p.22). Mearns (1997) quotes from clients and counsellors to help the reader understand

what is meant by relational depth, including, "I knew he knew I knew" (p.21) and "feeling me in the same moment…" (P.21), which suggest much more than a cognitive encounter. Mearns (1997) confirms this when he says, "…one of the ingredients involved in the therapist's ability to work at relational depth is a coming together of high levels of the therapeutic conditions of empathy, unconditional positive regard and congruence…" (p. 23).

It seems noteworthy to mention at this point, that this last quote by Mearns (1997) serves to distinguish relational depth from the concept of working alliance. Perhaps the most accepted definition of working alliance comes from Bordin's (1979) alliance model where agreement on the goals of therapy, agreement on tasks and the interpersonal bond make up the working alliance. Bordin also (1979) stressed the importance of the therapist and client liking each other, if they were to engage in long-term psychotherapy. In sharp contrast, Mearns (2003) states that liking is highly selective and is usually related to valuing only a certain aspect of a person. He stresses that unconditional positive regard, on the other hand, is more about valuing the whole of the person. He concludes that liking and unconditional positive regard are therefore not the same thing and that the two are not to be confused. In the same publication, Mearns (2003) implies that it is relational depth which makes person-centred counselling different by describing it as "an extraordinary depth of human contact" (p.5). However, a more comprehensive and more workable definition of relational depth comes from Mearns and Cooper (2005), who proposed the following definition:

A state of profound contact and engagement between two people, in which each person is fully real with the Other, and able to understand and value the Other's experiences at a high level' (p.xii).

It is clear that this definition has its roots in the person-centred approach to therapy and therefore is based on (albeit implicitly) the 'Necessary and Sufficient Conditions of Therapeutic Personality Change' as put forward by Rogers (1957). Mearns (in Mearns & Cooper, 2005) states that a relationally deep meeting can be seen as a single core condition where all of six of Rogers (1957) conditions are present and, what are known as the Core Conditions (3, 4 and 5), are present in high degree. The six conditions are:

1. Two persons are in psychological contact.
2. The first, whom we shall term the client, is in a state of incongruence, being vulnerable or anxious.
3. The second, whom we shall term the therapist, is congruent or integrated in the relationship.
4. The therapist experiences unconditional positive regard for the client.
5. The therapist experiences an empathic understanding of the client's internal frame of reference and endeavours to communicate this experience to the client.
6. The communication to the client of the therapist's empathic understanding and unconditional positive regard is to a minimal degree achieved.

Mearns and Coopers' definition (2005) implicitly captures the qualities that are explicitly covered by the six conditions of Rogers (1957). For example, the part of the relational depth definition that mentions profound contact reflects Condition 1 (two persons in psychological contact). The 'fully real' part of the definition not only reflects Condition 3, the congruent therapist, but also reflects client congruence. The 'understanding and valuing the other's experience' part of the definition reflects Condition 4's unconditional positive regard and Condition 5's empathic understanding. The part which says that "each person is…able to understand and value the Other's experiences" certainly captures Rogers' (1957) sixth and last condition. With regard to Condition 2 (an incongruent client), this may seem to be incompatible with the part of the relational depth definition that states "each person is fully real with the other." However, the client may experience relative incongruence on a general day-to-day basis when going about their daily lives, but also experience moments of congruence in the safe environment of a deeply therapeutic relationship with a competent therapist. In other words they may only experience moments of high congruence in their therapy sessions during moments of relational depth. Furthermore, the concept of relational depth does not specify that both client and therapist need to be in possession of the core conditions in equal measures. It may be enough that the therapist possesses them to high degree but to a higher degree than their client.

Although Rogers himself never formulated a concept such as *relational depth* which defined all of the conditions into a single core condition, he described many

experiences which are arguably close to describing the experience of relational depth.

> I find that when I am close to my inner, intuitive self, when I am somehow in touch with the unknown in me...whatever I do seems to be full of healing. Then, simply my presence is releasing and helpful to the other. There is nothing I can do to force this experience, but when I can relax and be close to the transcendental core of me...it seems that my inner spirit has reached out and touched the inner spirit of the other. Our relationship transcends itself and becomes a part of something larger. Profound growth and healing and energy are present (Rogers, 1980, p. 129)

Certainly the above description could easily be describing a 'self-experience' of relational depth especially from a therapist perspective and is possibly the closest Rogers came to in explaining a moment of deep encounter with a client.

Whilst focusing on this one definition, this project focuses on relational depth as a specific moment of encounter as opposed to focusing on relational depth as the particular quality of the relationship. That is not to say that one aspect is of more value than the other but rather to clarify that this research project's main focus is to research a moment, time and/or experience of relational depth and to create a measure that assesses such an experience.

2.1.2 A Review of Related Relational Constructs

2.1.2.1 Introduction

The following review looks at other concepts to relational depth that describe particular experiences in the relationship between client and therapist. The concepts

that are reviewed are those which bear some similarity to relational depth insofar as they name a particular relationship variable. The review looks firstly at the alliance which is by far the most documented and researched relationship variable between client and therapist. It then goes on to explain some other, lesser known, relationship variables that have something in common with the concept of relational depth. The reason for explaining and documenting such concepts is to put relational depth in perspective with other similar concepts.

2.1.2.2 The Alliance.

One of the most documented relational concepts in psychotherapy is that of the therapeutic alliance. Although a somewhat different concept to relational depth, the alliance can be seen as a relationship variable within therapy and as such it is relevant to the concept of relational depth. Whilst the two concepts are different, they are similar insofar as they both focus on aspects of the relationship between client and therapist. It is relevant therefore to look into detail as to what the alliance consists of in order that it may be contrasted to relational depth. Sometimes called the 'working alliance' (Greenson, 1965), 'helping alliance' (Luborsky, Crits-Christoph, Alexander, Margolis and Cohen, 1983) or 'therapeutic alliance' (Bibring, 1937) this concept of the relationship between therapist and client has its roots in psychoanalysis.

It was the psychodynamic school that first gave rise to the concept of a client-therapist relationship, even though the term alliance was not used until later. The psychodynamic school's original equivalent to a therapeutic alliance was transference. This transference was not a negative transference but more of a

positive transference that was seen by Freud (1958a) as 'divisible into transference of friendly or affectionate feelings…' (p.105). Such transferences were seen to be, to a degree, part of the unconscious and consisted of a process that enabled the client to have trust in the therapist (Freud, 1958a). The idea of positive transference, although different to the concept of relational depth, is comparable to it in that the friendly or affectionate feelings are not completely different from the intimacy that Mearns and Cooper (2005) talk about.

> Like the string in a game of cat's cradle, as the interrelating weaves its way around and between participants, so it draws them ever closer together. Other terms that can be used to describe the relationally deep encounter, then are 'connection', 'closeness', 'contact' and 'togetherness'. (p. 47)

Another psychodynamic theory concerning the alliance came from Sterba (1934) who stated that the analyst can affect an alliance with the analysand's ego by using one part of the ego. Thus, in a successful therapeutic relationship, the objective side of the client's ego enters into an alliance with the therapist. Similarly, Edward Bibring (1937) referred to the analyst and the healthy part of the patient's ego and first coined the phrase the *therapeutic alliance*. The idea of the client somehow entering into an alliance with the therapist is comparable to relational depth insofar as it is an expression of the relationship between therapist and client

During the 1950s, the concept of the therapeutic alliance became influential in the United States when Elisabeth Zetzel (1956) advocated its inclusion as a

psychoanalytic technique. Zetzel (1965) further developed this idea and saw the therapist's attitude similar to a mother toward their child; she also saw this therapeutic alliance as the condition for the emergence and resolution of the transference neurosis.

Stone (1961) described a 'physician-patient' relationship in psychoanalysis and postulated that there were three aspects to this: the real and actual (integrated) relationship, the transference-countertransference relationship and the activities that happen within psychoanalysis (such as techniques that include deprivations and prohibitions). Stone (1961) also stated here that the psychoanalytic setting reproduces the phases in a person's life, most notably the phases of separation from their mother.

The influence of psychodynamic theories continued with regard to the therapist-client relationship. Greenson (1965) added to Freud's work postulating that in the therapist-client relationship, there were three main elements; transference, a working alliance and the *real* relationship. Here, the real relationship referred to the 'true' relationship or what actually happened in the relationship; in other words this was, in a sense, what transference was not and was seen to be counter to transference. In the same publication, Greenson (1965) put forward the term *working alliance* and saw it as the client's ability to align with the tasks of analysis, differentiating with the *therapeutic alliance* which he saw as the ability to form a personal bond.

It was not until the 1970s that attempts were made to project and take forward the

concept of the alliance from its roots in the psychoanalytic school in order to encompass the relational elements in all theoretical approaches to therapy. Luborsky (1976) proposed an extension of Zetzel's (1956) and Stone's (1961) concept of the alliance. Having so many definitions of the alliance hindered research. A working definition was needed that embraced all orientations. Luborsky, an influential figure in the therapy research community, proposed a definition that attempted to encompass all orientations of therapy. Coming away from a single-approach model of the alliance, Luborsky (1976) was one of the first to put forward a pantheoretical definition of the therapeutic alliance. He proposed there were two types of alliance; Type 1 was a term given to describe the client's belief that the therapist was a source of help and support and Type 2 involved the client's investment, faith and sense of shared responsibility as they worked with the therapist towards a common goal. Later, Luborsky (1994) identified six signs that represented the Type 1 alliance and four signs which represented the Type 2 alliance. The four signs which represented the Type 1 alliance were 1) feeling the therapist to be warm and responsive, 2) believing the therapist was helping, 3) feeling changed by the treatment, 4) feeling a rapport with the therapist, 5) feeling the therapist respected him/her, and 6) conveying a belief in the treatment process. For the Type 2 alliance the four signs were the client 1) experiencing the relationship as working together in a joint effort, 2) sharing similar conceptions about the source of problems, 3) expressing beliefs about continually being able to cooperate with the therapist and 4) demonstrating abilities similar to those of the therapist in terms of being able to use the tools for understanding.

In probably his most noteworthy paper, Edward Bordin (Bordin, 1979) also stressed the need for a definition that encompassed all psychotherapy approaches. His mission was to bring about a coming together of therapeutic orientations with regard to defining the 'working alliance', and, as he put it "to call attention to a point of view that can encompass most, if not all, approaches to psychotherapy and can lead toward the needed convergence in research" (p.252). Although rooted in psychoanalytic theory, Bordin (1979) was proposing that the concept of working alliance could be generalised to all approaches. This was revolutionary at the time for it focused on the conscious aspects of the relationship rather than unconscious processes. It also focused on how therapist and client might work together and emphasized the collaboration between therapist and client. He also proposed that working alliance between the person who seeks change and the change agent "is one of the keys, if not *the* key, to the change process" (p.252).

Bordin (1979) described this relationship between client and practitioner as a 'therapeutic working alliance' (p.252). He went on to describe this alliance in terms of goals, tasks and bonds. With regard to the concept of goals, he described this specifically as more to do with agreement, where client and practitioner collaborate and agree on sometimes very specific targets in therapy. He stated that the specific goals were dependent on the type of therapy, but nonetheless he proposed that all therapy had goals. He described tasks in various ways depending on the type of therapy. For example, he explained that in psychoanalytic therapy the rule of free association is specified as a task, and in behaviour therapy a task might be to record the frequency of certain behaviours. Concerning client-centred therapy Bordin

(1979) said '…tasks are never explicitly specified, and only emerge gradually and ambiguously' (p.254). With regard to bonds, Bordin (1979) stated that if client and practitioner are proposing to work together for a number of years, then an important aspect might be whether they like or dislike each other. With regard to bond he also mentioned trust and, in long-term therapeutic relationships, emphasised the importance of "deeper bonds of trust and attachment" (p. 254).

Bordin's (1979) proposals in his 1979 paper, which was a milestone in the psychotherapy literature, have had an enduring quality. Some 20 or so years later, in a meta-analytic review, Martin, Garske and Davis (2000) stated that most definitions of therapeutic alliance have three themes in common: "(a) the collaborative nature of the relationship, (b) the affective bond between patient and therapist, and (c) the patient's and therapist's ability to agree on treatment goals and tasks" (p.2). However, Horvath and Bedi (2002) state that they do not have a universally accepted definition. They go on to suggest that there are important shared aspects with regard to definitions but there also appear to be differences in definition that are non-trivial. More recently, Horvath, Del Re, Flückiger and Symonds (2011), in a research synthesis of the relations between alliance and outcomes, come to the same conclusion.

In another pantheoretical approach to the alliance, Louise Gaston and Charles Marmar (Gaston and Marmar, 1994) proposed another definition. They proposed that the alliance was a combination of therapeutic alliance, working alliance, therapist understanding and involvement, as well as client-therapist agreement on

goals and strategies. Gaston and Marmar's (1994) proposed notion of therapeutic alliance had its underpinnings in Freud's (1958a) notion of transference as well as attachment to an identification with the therapist. Central to their idea of working alliance are both the client's and therapist's skills used in collaboration on the tasks of therapy. Gaston and Marmar (1994) argue that focus should not just be on the client but also on the therapist; the therapist's part in the alliance (their understanding and involvement) is included to indicate that therapists do not behave in a uniform way and that the therapist plays an important role in the development of the alliance. The aspect of client-therapist agreement on goals and strategies is based on Bordin's definition of the alliance of agreement on bonds, tasks and goals.

In sum, the concept of relational depth differs from the concept of the alliance; whilst relational depth emphasizes the mutual aspects of empathy, congruence and value at deep levels during a particular moment, the alliance largely emphasizes therapist and client collaboration on goals, tasks and bonds; whilst the experience of relational depth is seen as being at the forefront of therapy (perhaps in its more enduring form), the alliance is seen perhaps as more of a backdrop to therapy where it is utilized more as a vehicle to reach therapy rather than being central to therapy itself. However, there are many other concepts regarding the relationship that are more similar to the relational depth concept than the alliance concept is.

2.1.2.3 The Therapeutic relationship

Murphy (2010) states that the term 'therapeutic alliance' has been used synonymously and interchangeably with the term 'therapeutic relationship'. He further states that this is an unfortunate circumstance since the therapeutic alliance

refers only to those parts of the relationship which focus on collaboration of goals of therapy, agreement on how those goals are achieved and the bond between client and therapist whereas, the Rogerian term 'therapeutic relationship' refers to the person-centred facilitative conditions of empathic understanding, congruence and unconditional positive regard otherwise known as the 'core conditions' (Murphy, 2010). I would agree with Murphy's (2010) statement and consequently, I should note at this stage that from here on, unless otherwise stated, that I use the term 'alliance' to mean the part of the client-therapist relationship which comprises the goals, tasks and bonds mentioned above. Similarly, unless otherwise stated, I refer to the Person-centred therapeutic relationship as essentially consisting of the three facilitative conditions of empathic understanding, congruence and unconditional positive regard. The reason for referring to the therapeutic relationship in this way is because of Rogers (1957; 1959) concept of the therapeutic relationship.

Rogers (1957; 1959) saw the effective therapeutic relationship as consisting of the therapist having a deep respect (or an unconditional positive regard) for the client. In addition, the therapist should be genuine or congruent and should have an empathic understanding for the client. If the client percieved the congruence, empathic understanding and unconditional positive regard then, according to Rogers (1957;1959) therapeutic movement would occur. Rogers (1957) believed that if a person was offered these conditions then they experienced a psychologically healthy environment. In this environment Rogers (1957; 1959) stated that the person would be able to know themselves better in that the person's self-concept will be more consistent with their organismic self.

Later in his career, Rogers (1980) put forward what could be termed a fourth condition; that of presence. When describing the quality of presence Rogers (1980) talked about being close to an inner intuitive self and being in touch with the unknown in him. He also talked about being in a slightly altered state of consciousness.

Therefore, the therapeutic relationship is conceptually different to the alliance and more similar to the concept of relational depth. Indeed, the therapeutic relationship is conceptually very similar to relational depth because both have the Rogerian 'necessary and sufficient' conditions as their theoretical underpinnings. Consequently, both theoretically consist of a combination of core conditions. However, the key differences between the two are that the therapeutic relationship refers to the general continual ongoing qualities of empathic understanding, congruence and unconditional positive regard that is provided by the therapist to the client, whilst relational depth refers to the mutual experiencing of such qualities; both therapist and client each providing such qualities to the other. Furthermore, unlike the therapeutic relationship, the experience of relational depth can be momentary and does not necessarily refer to the continual ongoing client-therapist relationship (Mearns & Cooper, 2005). Therefore, whilst an experience of relational depth cantheoretically occur within a therapeutic relationship, it is not the whole therapeutic relationship itself, but rather something that may, or may not, occur within it.

2.1.2.4 Martin Buber and the I-Thou concept

Buber (1958) put forward the concept of the 'I-thou' attitude and distinguished it from something he called the 'I-It' attitude. The two terms, 'I-Thou' and 'I-It', are described as primary word combinations in that although each of the terms consist of two words, they are each seen to be one word; these two primary words are not isolated words but rather words that have been combined to make one word. Both these primary word combinations, I-It and I-Thou, refer to experiences towards the other rather than a particular kind of relationship. It is the I-Thou encounter that bears most resemblance to the relational depth concept. However, in order to understand the I-thou, one needs to also understand the I-It.

Buber (1958) postulates that the I-It encounter refers to the experience that the self might have with something or somebody that the self objectifies with. It refers to the self that is the self in relation to the it. This means that the 'It' in I-It could refer to a person, as in 'I-He' or 'I-She' but that if it did, it would be because the self was objectifying that person. Therefore it can be seen that the *I-It* experience is a situation where the self encounters something or someone that is objectified by the self, where that something or someone can be experienced and utilized by the self. Buber (1958) saw that we could experience things and people in the world whilst objectifying them and that when the self experiences things and people in this way; there is an I-It experience.

In sharp contrast, Buber (1958) stated, 'The primary word *I-Thou* establishes the world of relation' (p.21). In other words, an *I-Thou* stance refers to the relationship that happens when two people form a connection; when there is a joining of two identities in an authentic, meaningful way. According to Buber (1958), when there is

an I-Thou situation, each person does not merely experience the other, rather each stands in relation to the other. Further, neither the *I* alone, nor the *Thou* alone, can experience this relation; the relation happens *between* the *I* and the *Thou.*

In an I-Thou position, the *I* is shared and the *Thou* is accepted. In I-Thou experiences, meaning occurs because there are no barriers and no facades; such an absence of barriers and facades enables authentic relationship (Buber, 1958). In this vein, by being in relationship with another, we are able to be in a relationship with ourselves and that gives us meaning according to Buber (1958). *I-Thou* positions are not about experiences per se; they are not necessarily about simply giving to another or having our emotional needs met by another. I-Thou experiences are about truly meeting another by stepping out of oneself; by coming away from mere experience and by truly being in relationship with another.

In light of the above, the *I-Thou* concept has much in common with the relational depth concept in that there is an absence of barriers and facades. This of course is comparable with the 'each person is fully real with the other' part of the relational depth definition. Another similarity the I-Thou concept has with that of relational depth is the fact that in I-Thou encounters, the I is shared and the Thou accepted. Mearns and Cooper (2005) propose that the therapist's experience of relational depth involves the therapist being fully transparent as well as possessing high amounts of acceptance toward their client. In this light, the therapist is seen to be sharing themselves and accepting the other. However, unlike the concept of relational depth, the I-Thou concept does not distinguish between specific moments of connection and

more enduring relationships. In addition, an I-thou attitude is really a position that one has towards another rather than it being a particular type of relating.

2.1.2.5 Linking

John Rowan (1998) describes the term linking as 'a special kind of empathy, a special kind of counter transference and/or a special kind of identification' (p.245). Essentially he describes linking as occurring at the *Subtle level* as opposed to the *Causal level* of the transpersonal self. According to Rowan (1998) the transpersonal refers to a spiritual tradition stating that it is not the extrapersonal, not the New Age, not the right brain and not religion (Rowan 1998) and the two levels of the Subtle and the Causal are levels of the transpersonal. Where the Causal level is characterised by typically no interest in symbols, no interest in gender, and does not distinguish between self and other, the Subtle level is completely the opposite where it is fascinated with symbols, concerned with gender and experiences deep linking with the other. This Subtle level can be thought of as where 'the intuition of the therapist comes fully into play, and the boundaries between therapist and client disappear' (Rowan, 2010, p.129). In describing linking more fully Budgell (1995) gives a comprehensive description.

> The experience is described as near fusion, a communion of souls or spirits, and a blurring of personal boundaries. To achieve this both parties have to give up something of themselves while remaining separate.
>
> It is not symbiosis, but the other end of the spectrum, as described by Wilber (1980). It is the transpersonal sense of relinquishing self. Symbiosis is about being cozy, but this is about working through pain and fear.

> It is a sacred experience and yet natural, and there all the time. It comes from the spiritual or transpersonal realm, being a step beyond empathy and the natural plain (p. 33).

With regard to seeing linking as empathy, Rowan (1998) goes on to say that from Budgell's (1995) description, linking is a special kind of empathy. He explains that empathy usually has been seen to have an 'as if' quality. In other words, in an act of empathy, we can imagine what it might be like for somebody but also know that we are completely separate from them. However, from this perspective, the concept of empathy as a form of linking is different to this 'as if' quality but rather it is seen to be an experience where two worlds overlap. Rogers (1980) is cited to explain this phenomenon further; 'It seems that my inner spirit has reached out and touched the inner spirit of the other. Our relationship transcends itself and becomes part of something larger' (p.129). Rowan (1998) also cites Mearns (1996);

> Similarly the therapist who is willing to work with the client at relational depth tries to leave outside conventional ways of responding and projects himself or herself fully into the client's experiencing. (Mearns, 1996, p.310)

Further, Rowan (1998) refers to Hart (1997) and describes linking as a typical Subtle-level phenomenon that can also be described as *Transcendental Empathy* (a term coined by Hart). Transcendental Empathy, as described by Hart (1997), is a more direct knowing of the client's world and thus has less of an 'as if' quality. Hart (1997) states that all empathy has the potential to be transcendental in that it has the

potential to change as well as being able to take both therapist and client out of themselves as well beyond a self-separateness. Hart (1997) also states that this self-separateness is a mutuality that has sufficient potency to be able to not only help the client develop, but help the therapist develop too.

In this sense, this is describing transcendental empathy as a kind of knowing, but not a knowing in the conventional sense, but rather a knowing that goes beyond the rational-empirical knowing and beyond the subject-object (or what Buber (1958) might call I-It). This consequently means it is a more immediate knowing of the other person.

The concept of linking is also seen to be special kind of countertransference, which can be accessed in a conscious way (Rowan, 1998). This is unlike the conventional psychoanalysis school (e.g. Freud, 1958b) that saw countertransference as something which only belonged to the unconscious. In this manner, the psychoanalytic concept of identification is also seen as a type of linking.

Again, within the psychodynamic school, identification is something that is conventionally seen to be as something in the unconscious (Freud, 1958b). However, from the linking perspective, identification is seen to be something that is accessed in a conscious way (Rowan, 1998). Not only is linking seen to be a conscious state but Rowan (1998) suggests it can be under the full control of the participant. In this light Rowan refers to Alvin Mahrer (Mahrer, 1989, cited in Rowan 1998) who practices something called *experiential listening* where the therapist reproduces in their own body whatever is happening to the client. Manuals

and workshops exist on how one does 'experiential listening.'

Later, Rowan (2005) describes linking as an experience arguably more similar to that of relational depth in that it is seen to be a moment of deep connection. Rowan (2005) suggests that such moments are characterized by our defences being lowered where '…we can genuinely let go of our boundaries and be with another person's soul…' (p. 111).

Essentially, it would appear that the concept of linking has much in common with the relational depth concept in that it is seen to be something which involves a deep connection experienced in the moment. The essential difference between the two concepts is that linking can be seen to be a both a spontaneous experience and/or a fully controllable experience. This is in sharp contrast to Mearns and Cooper's (2005) view of relational depth where they state that relational depth can only be facilitated, not created or made; in other words, theoretically, relational depth is not a fully controllable experience.

2.1.2.6 Real Relationship

An important construct comparable to relational depth is that of the *real relationship* as put forward by Greenson (1965) and later Gelso (2011). Here, the real relationship refers to that relationship that exists between therapist and client that is the transference-free element. Gelso (2011) views the real relationship as consisting of two elements a) genuineness and b) realism and describes it as "the personal relationship existing between two or more persons as reflected in the degree to which each is genuine with the other and perceives the other in ways that befit the other"

(p.12-13). Genuineness is viewed as being authentic and opposed to being fake or phony. The realism aspect is the experience of perceiving the other person in a way that befits that other rather than what the perceiver wishes for needs or fears.

However, the concept of the real relationship is different in many ways to relational depth. In the present study relational depth refers to moments, events and relatively short-lived experiences but Gelso (2011) refers to the real relationship as showing itself in the thoughts of participants outside of the therapy hour. Another contrast is that the components of the real relationship do not seem to include any form of valuing of the other's experiences. Although clearly related as a concept to relational depth, the real relationship refers generally to accurately presented and perceived aspects of the therapeutic relationship, as opposed to the transference. As such, relational depth can be understood as a particular kind of real relationship, that is, one in which client and therapist value each other positively and experience a level of intense or deep psychological contact.

2.1.2.7 Moments of Meeting

Daniel Stern (2004) puts forward the idea that therapy 'moves along' in that it is the everyday dialogues that moves therapy forward. Elements that make up this moving along include those *present moments* where one is merely aware and those present moments where one is fully conscious. Therefore, it is postulated that the former do not enter long-term memory but in contrast the latter, that is the conscious present moment, is seen as a first-person phenomenon open to introspection (and therefore capable of entering long-term memory). Conscious present moments are seen as having three different types; 1) regular present moments, 2) now moments and 3)

moments of meeting. Regular present moments are seen as essentially everyday occurrences which happen during therapy. Now moments are seen as those moments which suddenly pop-up and are highly charged with impending consequences; they are what is termed as 'heavy with presentness' and the need to act. According to this theory, moments of meeting usually follow now moments.

Stern (2004) goes on to describe moments of meeting as consisting of two people experiencing an inter-subjective encounter where each party is aware of what the other is experiencing. In Stern's (2004) words a moment of meeting is

> a present moment in which the two parties achieve an intersubjective meeting. At this moment they are aware of what the other is experiencing. They share a sufficiently similar mental landscape so that so that a sense of "specific fittedness" is achieved (p. 151)

Stern (2004) goes on to further describe moments of meeting as a *shared feeling voyage* (p. 172) where 'two people traverse together along a feeling landscape as it unfolds in real time (p. 172). It is easy to see the comparison between relational depth and moments of meeting, especially the relational depth that describes specific moments as opposed to enduring experiences. Indeed, a specific moment of relational depth could be seen to be a moment of meeting. However, unlike relational depth the theory behind a moment of meeting does not explicitly address the valuing that each party in the client-therapist relationship have for each other.

2.1.2.8 Sacred Mirroring: Being Together

In the practice of pastoral counselling, *therapeutic mirroring* refers to the therapist having the capacity to empathise with the client but in such a way that the therapist reflects back the essential nature of the client (Pembroke, 2006). Pembroke further puts forward that mirroring is an act of love (Pembroke, 2006). Therapeutic mirroring comes from a *nonduality* philosophy where nonduality simply means 'not-two' and derived from the Sanskrit word *advaita* (meaning not two) (Prendergast, Fenner & Krystal, 2003). Essentially, it sees the split between self and other as being merely a mental construct (Prendergast, Fenner & Krystal, 2003). Central to the idea of *sacred mirroring* is an inherent healing awareness that unfolds between and transcends the therapist and client (Prendergast, Fenner & Krystal, 2003). Normal mirroring is therefore seen to be reflecting the client's own personal experience. However, sacred mirroring 'reflects back the impersonal Ground of Being shared by client and therapist' (Prendergast, 2003, Fenner & Krystal, p. 95). In this sense sacred mirroring can refer to the intimacy that arises from client and therapist. However, it is seen to be an impersonal event in that the line between therapist and client becomes blurred. Neither client not therapist can be a sacred mirror or be sacredly mirrored; it is simply by being that sacred mirroring occurs. The *being together* part of sacred mirroring is to do with a shared being with what arises and is an intimate and quiet kind of contact (Prendergast, Fenner & Krystak 2003). Prendergast, Fenner & Krystal (2003) report that clients have reported sensations of dropping, opening or deepening into a more expansive awareness.

It can be seen that the concept of sacred mirroring has some commonality with relational depth in that both client and therapist are on a shared plain.

2.1.2.9 Third Space or Analytical Third

Other intersubjective or relational models (also within psychodynamic approaches), put forward the idea of a "third space" or "analytical third" (Ogden 1994a). Here the analytic process is seen to reflect three subjectivities: the first is the therapist, the second is the client, and the third, that is the analytic third, is seen to be a creation of therapist and client. In addition, both therapist and client are also created by the analytic third. Ogden (1994a) states that in spite of the analytic being a creation of therapist and client, it is not seen in the same way by each party due to their own personality system, personal history and other aspects unique to that person.

Furthermore, the analytic third is seen to be asymmetrical in construction due to its creation being in the therapy setting where it is defined by the relationship of client and therapist (Ogden, 1994b). The analytic third is seen as the intersubjectivity of the therapist-client relationship as it is generated by and between the separate subjectivities of the therapist and client and is consequently seen to be a product of a unique dialectic (Ogden, 2004). As a consequence, the client's unconscious experience is seen to be privileged in that the client's past and present are, in a sense, taken up by the therapy dyad and are the central topic of discourse (Ogden, 2004). In light of this and from the therapist's perspective, the analytic third can be seen to be a vehicle used to understand the client's conscious and unconscious experience.

Moodley and Lijtmaer (2007) later comment on the analytic third concept in relation to self-disclosure in therapy stating that those with a cross-cultural analytic stance tend to support the idea of an analytic third. Furthermore, those with such a stance see the analytic third as a place where disclosure is possible and which is 'beyond

technique and skill and supports the development of empathy, compassion and relational interaction' (p.49).

The idea of an analytic third is not so far removed from the notion of relational depth where Mearns and Cooper (2005) indicate that in order to facilitate a meeting at relational depth, the therapist needs to let go of techniques, aims and anticipations. Furthermore, the idea of a space where disclosure is possible is similar to the findings of Knox (2008) where clients' experiences of a moment of relational depth was typically characterized by feelings of 'being on a different level' and of feeling the 'need to verbalise/bring out/give voice to parts.'

2.1.2.10 Tenderness

Brian Thorne (1991) proposes that when a person possesses the quality of 'tenderness' in all its fullness it means, firstly, that the quality will irradiate the total person and that it will be evident in voice, eyes, hands, thoughts, feelings, beliefs, moral stance. Secondly, it communicates through it responsiveness. Thirdly, it demonstrates a preparedness and an

> ability on the part of the therapist to move between the worlds of the physical, the emotional, the cognitive and the mystical without strain and by a willingness to accept and celebrate the desire to love and to be loved if and when it appears in the therapeutic relationship. (Thorne, 1991, p. 76)

Thorne (1991) indicates that if we can develop the quality of tenderness in ourselves then we can have the hope that there will indeed be a future and that it will be a time

in which something qualitatively different can happen between two human beings (p.77). He further adds that 'when tenderness is present in a relationship I believe that there is the possibility of finding wholeness and of recognizing the liberating paradox' (p.77). Thorne (1991) also states that the quality of tenderness is fleeting and momentary and that during such moments client and therapist can be caught up in a 'stream of love.' Knox (2011) gives a breakdown of the qualities of tenderness that Thorne (1991) describes

1. A sense of heightened awareness.
2. Feel in touch with myself.
3. It is as if energy is flowing through me.
4. I feel a physical vibrancy, and this often has a sexual component.
5. I feel powerful and yet at the same time almost irrelevant.
6. My client seems more accurately in focus.
7. When he or she speaks, the words belong uniquely to him or her.
8. Physical movements are a further confirmation of uniqueness.
9. It seems as if for a space . . . two human beings are fully alive.
10. My client and I are caught up in a stream of love . . . within this stream. there is an effortless or intuitive understanding.
11. A sense of joy which, when I have checked it out, has always been shared by the client.
12. Tears may flow or there may be a sudden outburst of laughter.
13. There may be an urgent need to talk about God or the Soul.
14. There may be an overwhelming desire for physical contact.
15. There is always a sense of well-being, of it being good to be alive.
16. I receive my client whole and thereafter posses a knowledge of him or her which does not depend on biographical data (p.18).

In light of the above, the quality of tenderness and the state of relational depth seem to be very closely linked.

2.1.3 Empirical evidence on the existence of relational depth.

2.1.3.1 Introduction to 2.1.3

Before any research is conducted to create a measure designed to assess an experience of relational depth, it is important to find evidence as to whether

relational depth actually exists. Whilst it may make perfect sense intuitively, that such an experience does exist, such intuition is not evidence enough if one is going to invest time in creating an inventory designed to assess such a concept. Fortunately, just before embarking on the current research, there were in existence a few studies that provided such valuable evidence. This subsection explores what evidence there is that relational depth exists and if so what the experience is like and whether such an experience has any impact on the client.

2.1.3.2 Research Carried Out by Current Author

The findings of the MSc study (Wiggins, 2007), carried out by the current author, essentially found that relational depth appears to consist of various experiences including, but limited to, connectedness, mutuality, a heightened self, an immersed self as well as feelings of trust, empathy (towards other and from other) and unconditional positive regard (towards and from other). Please see section 1.3 for more information on this MSc research.

2.1.3.3 Research Carried Out by Mick Cooper

Cooper (2005) initiated research on relational depth by interviewing therapists about specific times in which they felt they had experienced relational depth. Results indicated that therapists experienced moments of relational depth with their clients and that there were many commonalities amongst therapists' descriptions. Such commonalities included heightened feelings of empathy, acceptance and receptivity towards clients; powerful feelings of immersion in therapeutic work; being real; increased perceptual clarity; and greater levels of awareness, aliveness, openness and satisfaction. Therapists also reported that during such moments of relational depth

they experienced their clients as highly transparent; articulating core concerns and issues; and reciprocating the therapists' acknowledgement of them in a flowing bi-directional encounter. Here, Cooper (2005) proposes that such experiences and descriptions of relational depth can be categorised into a tripartite system consisting of self-experiences, perceptions of client and experiencing the relationship. This research was groundbreaking in that it uncovered the characteristics of relational depth between therapist and client. However, it only does so from a counsellor's perspective.

2.1.3.4 Research Carried out by Rosanne Knox

To shed more light on clients' experiences of specific moments of relational depth, Knox (2008) focused on clients' experiences of the moment; then later Knox and Cooper (2010), with a completely different sample, focused on the relationship qualities associated with a moments of relational depth. In both of these studies a phenomenological approach was used where clients were interviewed using semi-structured interviews. In the former study, 14 therapist-clients (clients who were also therapists or trainee therapists) were interviewed and in the latter 14 'bona-fide' clients (non-therapist client who were not even trainee therapists) were interviewed. In both studies clients reported that they identified one or more experiences of a moment of relational depth with at least one therapist. The experience of the moment itself was described as in another dimension, with a sense of spirituality, healing and empowerment.

In the former study Knox (2008) found that clients' experiences of a relational depth moment were that they felt real, open, deeply understood and wholly accepted.

Clients experienced their therapist as open, holding, accepting, being real and offering something over and above what they had expected from a professional therapeutic relationship. In the latter study (Knox & Cooper, 2010), the perception of the client-therapist relationship was emotionally close with an 'understanding beyond words'.

In both studies, clients described the moments as a catalyst, and saw them as highly significant moments in therapy, and as having a positive effect on the therapeutic process and on their lives after the therapy had ended. Another finding was that during relational depth, participants viewed themselves as being a 'proactive agent.' The non-therapist clients also emphasised the role of their own perseverance throughout the relationship.

Knox's (2008) first study results bore similarity to Cooper's (2005) where a similar category structure emerged; categories comprised experiences of self, of therapist and of relationship. However, a key difference in Knox's study was that clients reported feeling like they were being invited in by their therapist. This contrasts with Cooper's (2005) study with therapists where there was no indication that therapists were inviting the client in any way. Another key finding of Knox's (2008) study was that nearly all clients in her sample reported having felt like they were being emotionally held in the specific moment. In Cooper's (2005) study there was no indication of this. Perhaps the most significant finding of Knox's (2008) study was surrounding the positive impact and positive enduring effects of the specific moment of relational depth. Knox's research provides initial evidence that clients perceive a

moment of relational depth as making a positive contribution to outcome. This would suggest that further research into relational depth and outcome would shed more light on to what extent a moment of relational depth would contribute to outome.

2.1.3.5 Research Carried out by McMillan and McLeod

Also, addressing the need for research into clients' experiences of relational depth, McMillan and McLeod (2006) focused on clients' experiences of relational depth within the therapeutic relationship as a whole (as opposed to specific times, experiences or specific moments) by interviewing 10 therapists who drew on their experiences from having been a client. They then analysed interview data using a system of open coding as well as consultation with colleagues to enable a consensus of opinion concerning the categorisation of descriptions. Their results concluded that 'letting go' was an important aspect of a deeply therapeutic relationship and consequently this emerged as a core category in their analysis. With regard to the term 'letting go', McMillan and McLeod (2006) here state that they found significant evidence that in a deeply therapeutic relationship the decision to let go can almost be like a 'leap of faith'. 'Experiencing a deeply therapeutic relationship' and 'Experiencing an inadequate therapeutic relationship' were key subcategories. In the former, they found that clients' experiences of a therapeutic relationship were concerned with 'being ready to engage with the therapist', and in the latter they found that clients experiencing wanting 'to get out of the relationship' also reported feelings of anger toward the therapist. The categorisation of their data is very different to Cooper's (2005) categorisation of therapists' descriptions suggesting that clients' experiences of relational depth are different from therapists'. It could also

suggest that the enduring quality of relational depth is experienced quite differently to specific moments or times of relational depth.

2.2 Previous Measures of the client-therapist relationship

2.2.1 Introduction

Many researchers have created questionnaires and/or measures that are designed to assess various relationship variables including alliance measures. There are numerous alliance scales used by researchers. Horvath and Bedi (2002) state that there are more than 24 scales currently in use but that four are used in the majority of research. To a degree, I have followed their lead and offer a short review of these four sets of measures. However, I also include reviews of other measures which are arguably more closely related to the concept of relational depth. These include the Barrett-Lennard Relationship Inventory, The Real Relationship Scale and the Helping Alliance Scale.

2.2.2 Working Alliance Inventory

The Working Alliance Inventory (WAI) was developed by Adam Horvath and Leslie Greenberg (Horvath & Greenberg, 1989). Essentially, this is a questionnaire designed to measure Bordin's concept of alliance, which consists of three factors; agreement on goals, agreement on tasks and personal bond with the therapist (cf. Bordin, 1979, 1994). Consequently the WAI has three subscales of goals, tasks and bonds with each one having an equal number of items (12). The original instrument had 36 items and a 7-point Likert scale (from 1= never, to 7=always). Therapist, client and observer versions were created.

The WAI has been found to have good convergent-related validity with high correlations with the global California Psychotherapy Alliance Scale (CALPAS) and the subscales of Goals (.84), Tasks (.79) and Bond (.72) (Horvath, 1994). Evidence of discriminant validity is a little more complex. Horvath (1994) states that most studies which test the WAI for discriminant validity compare the relation of the WAI and other alliance measures with the relation of the WAI and the Counselor Rating Form (CRF; LaCrosse, 1980) which is designed to assess expertness, attractiveness and trustworthiness. Results have found that the WAI's relation with the CRF is significantly lower than the WAI's relation with other alliance inventories (Horvath, 1994) evidencing discriminant validity. The WAI has also been found to have excellent reliability, with internal reliability coefficients ranging from .93 to .84 for the global scale of the WAI. The WAI has also been shown to be significantly related to outcome with correlation coefficients ranging from .11 to .87 with the average being .33 (Horvath, 1994).

Shorter versions have been created with the Working Alliance Short Form (WAI-S) (Tracey & Kokotovic, 1989) and more recently, the Working Alliance Short Form-Revised (WAI-SR) (Hatcher & Gillaspy, 2006). The WAI-SR was tested using two samples in order that cross validation analyses could be conducted. In this latest short form, Hatcher and Gillaspy (2006) found that a five-point Likert scale improved the inventory as they found that respondents discriminated more effectively especially in the lower ends of the scale. They also state that the Goal, Task and Bond dimensions were better discriminated than in past versions. This revised version was found to yield excellent consistency with the older version

yielding coefficient alphas from .85 to .90 for subscales, and .91 and .92 (Sample 1 and Sample 2 respectively) for total score alphas. The WAI was found to correlate well with other alliance measures in that it correlated highly with the CALPAS (correlation coefficient .80) and with the Helping Alliance Questionnaire (HAQ) (correlation coefficient .74). The Goal, Task and Bond subscale correlations with the CALPAS were .63, .79 and .65 respectively and with the HAQ were .57, .74 and .59 respectively.

2.2.3 Penn Helping Alliance Scales

One of the first sets of tools to measure the therapeutic alliance were the Pennsylvania Helping Alliance Scales (Luborsky, Crits-Christoph, Alexander, Margolis & Cohen, 1983), which were essentially designed to measure the warmth and support that the client experienced in the therapist as well as the level of collaboration also experienced by the client. Also known as the Penn Scales, these inventories were created to test Luborsky's (1994) psychodynamic concept of the alliance in that they were designed to assess Type I and Type II forms of alliance. Type I was designed to measure the patient's experience with regard to how helpful the therapist was and Type II was designed to measure the patient's experience of collaboration with the therapist towards goals. The scale consists of 10 items and is the shortest of the alliance measures. Six items measure the client's experience of receiving help, or a helpful attitude from the therapist, and four items measure the client's experience of collaboration with the therapist. The items are rated on a 10-point Likert scale. Interrater reliability has ranged from .75 to .88 for the scales and

internal reliability has been reported at .91 for the total measure (Morgan, Luborsky, Crits-Christoph, Curtis, & Solomon, 1982).

A more recent version of the Penn Scales is the Helping Alliance Questionnaire (HAq), with the latest revised version being the Helping Alliance Questionnaire II (HAqII) (Luborsky et al, 1996). This measure is still intended to measure Types I and II forms of alliance as described above, but is designed to address construct validity problems by removing items which assessed early symptomatic improvement and replacing these with items focusing more on the alliance. The scale consists of 19 items with patient and therapist versions. Luborsky et al (1996) found that the HAqII has excellent internal consistency in both client and therapist versions (.90 and .93 respectively) and excellent test-retest reliability (.78 for the client and .56 for the therapist version). The authors also found that it had good convergent validity with the California Psychotherapy Alliance Scale (CALPAS) in that correlations between the patient version of the HAqII and CALPAS subscales ranged from .38 to .71. The therapist versions of the HAqII and CALPAS were slightly higher, ranging from .61 to .79.

2.2.4 Vanderbilt Scales

The Vanderbilt Psychotherapy Process Scale (VPPS), which was developed by O'Malley, Suh and Strupp (e.g. O'Malley, Suh & Strupp, 1983; Suh, Strupp & O'Malley, 1986) at Vanderbilt University, was designed to measure the therapist-patient relationship and more specifically the psychotherapy process. This scale was designed so that a given segment of therapy could be observed and rated, and was specifically created to assess dynamic therapies. Therefore, this instrument is an

observer-rated scale rather than being rated by the participants of therapy (i.e. client-rated or therapist rated). More specifically, it was designed to focus on positive and negative aspects of both the client's and therapist's behaviour which may have helped or hindered therapy. It consists of 80 items each rated on a 5-point Likert scale that are designed to be unidimensional, descriptive and requiring very little inference. After a number of refinements, the VPPS resulted in eight stable factors: 1) patient participation, 2) patient hostility (resistance) 3) patient exploration, 4) patient psychic distress, 5) patient dependency 6) therapist warmth and friendliness, 7) therapist negative attitude and 8) therapist exploration.

As this was not designed specifically to measure the alliance, the Vanderbilt Therapeutic Alliance Scale (VTAS) (Hartley & Strupp, 1983) was later created. A total of 44 items were created and rated on a 6-point scale where 14 items are relevant to patient, 18 to the therapist and 12 to the therapist-patient interaction. Factor analysis has shown it to yield six factors 1) positive climate, 2) client resistance, 3) Therapist intrusiveness, 4) client motivation, 5) client responsibility and 6) client anxiety. This provides evidence that it assesses aspects of the alliance and thus is an improvement on the earlier measure. Reliability coefficients have ranged from .74 to .90 for the VTAS. This newer instrument has now replaced the earlier instrument at Vanderbilt.

2.2.5 California-Toronto Scales

The California-Toronto Scales were essentially a 'family' of inventories originally developed by Marmar and colleagues (e.g. Marmar, Weiss, & Gaston, 1989). These scales appear to be again influenced by the more dynamic therapies. Many versions

of these scales have been designed and revised over time and have branched out into several different scales.

The Therapeutic Alliance Rating Scale (TARS; Marziali, 1984) was developed and designed to assess therapist and client contributions to the therapeutic relationship. Three versions were created; patient, therapist and judge (observer). Each version of the alliance measure consisted of 21 items (11 positive and 10 negative) assessing the therapist contribution and 21 items (11 positive and 10 negative) assessing the patient contributions. Four subscales made up the 21 items; 1) patient positive contribution scale, 2) patient negative contribution subscale, 3) therapist positive contribution subscale and 4) therapist negative contribution subscale. Marziali (1984) found that the TARS had good internal reliability for all three versions with Alphas ranging from .81 to.93. In the subscales, both patients' and therapists' positive contributions and negative contributions, were highly correlated as judged by each member of the therapy dyad (Alphas ranged from .70 to.82). With regard to outcome, the TARS patient-rated subscales of patient positive contribution and therapist negative contribution to the relationship, most strongly predicted outcome after pretherapy was controlled for (with partial correlations of .38 and .47 respectively). No other subscale, regardless of vantage point, positively or negatively significantly predicted outcome.

The California Psychotherapy Alliance Scale (CALPAS) began as a 41-item measure designed to be rated by independent judges. Marmar et al (1989) revised the measure and created the original version of the CALPAS which contained 31 items

for the patient version and 5 items for the therapist and rater versions. Later a 24-item measure designed to assess four different dimensions: patient working capacity, patient commitment, therapist understanding and working strategy consensus was created and remains the most recent version (Gaston & Marmar, 1994). The CALPAS scale is an observer rated measure where raters make inferences about complex clinical concepts. As a result raters are usually trained to develop rating skills and are also therapists. With regard to validity, the CALPAS has been found to have significant correlations with other similar measures such as the Penn (.62), the VTAS (.38) and the WAI (.37) (Fenton, Cecero, Nich, Frankforter & Carroll, 2001). It has been shown to be significantly related to outcome with correlations of .63 for Cognitive Behaviour Therapy and .42 for the 12- step approach (Fenton et al, 2001). Regarding reliability, intraclass correlations of .76 were found for interrater reliability.

2.2.6 Real Relationship Inventory.

The Real Relationship Scale and Inventory were both created and designed to assess the psychodynamic construct of the real relationship as put forward by Gelso and Carter (1994) and later Gelso (2011) (see Section 2.2.4 for an explanation of the Real Relationship). Essentially therefore these two measures are designed to assess the transference-free aspects of the client-therapist relationship. The Real Relationship Scale was developed by Eugster and Wampold (1996) as one part of a battery of measures in order to evaluate treatment success. The scale is designed in such a way that both client and therapist make ratings of the quality of the real relationship as offered by the other person in the dyad. Each version consists of 8 items each having 6 anchor points. This meant that patients would be given two inventories; one

assessing their own real relationship and one to assess the therapist's real relationship. Consequently, four sets of scores would be collected for each therapy dyad assessed.

In the former, an example item is 'I did not like my therapist in these sessions' and in the latter, the corresponding item is 'In these sessions, my therapist really cared about me as a person.' Gelso (2011) states that although these measures were tailored to meet the theoretical formulations of Gelso and Carter (1994), closer inspection of the self items (the client's rating of themselves) reveal that these items are actually assessing the client's liking and caring for the therapist. Gelso (2011) implies that the Real Relationship Scales were not completely valid in that they did not focus on the realism construct but instead assessed only genuineness as well as a liking and caring for the other.

Eventually a more valid and reliable Real Relationship Inventory was created. From an initial item pool of 130, Gelso (2011) eventually created the 24-item Real Relationship Inventory. Items were created by a team of academics who were three doctoral-level psychologists with post-doctoral experience and three doctoral students. Before any items had been created, a decision was made that the final inventory would have 24 items. The final 24-item inventory consists of two subscales; 12 items designed to assess Genuineness and 12 items designed to assess Realism. This first version of The Real Relationship Inventory – Therapist (RRI-T) was designed to be completed by therapists but the focal point of the item could be the therapist, the client or their relationship. After administration to a development

sample, data showed that the internal consistency of the 24 items yielded alpha coefficients of .89, .87 and .93 for the Realism subscale, the Genuineness subscale and the total score respectively. With regard to validity, the RRI-T is significantly, positively correlated to the Working Alliance Inventory, the Depth and Smoothness subscales of the Session Evaluation Questionnaire as well as Emotional and Intellectual Insight, demonstrating convergent validity. It is also shown to be negatively associated with Negative Transference and uncorrelated with Positive Transference and Social Desirability thus demonstrating discriminant validity (Gelso, 2011).

Gelso (2011) created The Real Relationship Inventory-Client (RRI-C) in the same way as the therapist version described above. It was then administered to 94 participants who were clients (e.g. clients recruited from GP surgeries). Internal consistency of the 24 items was excellent at .90 for the Realism subscale, 91 for the Genuineness subscale and .95 for the total score. Regarding validity, the RRI-C strongly correlates with the Working Alliance Inventory (.79), Therapist Real Relationship (.60) and the Therapist genuineness scale of the Barratt-Lennard Relationship Inventory (.71) each demonstrating convergent validity. The RRI-C also demonstrated good discriminant validity that it is negatively related to other-directedness (-.24) (on the assumption that being genuine is being self-directed and not hiding one's feelings or being directed by others).

Lastly, the RRI-C has been found to successfully predict therapy outcome. Fuertes et al (2007, cited in Gelso, 2011) found that both the RRI-C and RRI- T significantly

predicted outcome (r =.36 and r = .49 respectively). Furthermore, despite the fact that both therapist and client versions of the RRI and the WAI were each separately associated with outcome, the RRI-C predicted outcome over and above that accounted for by the WAI. However, the 5% increase did not quite reach statistical significance.

2.2.7 Barrett-Lennard Relationship Inventory

Barrett-Lennard Relationship Inventory (BLRI) (Barrett-Lennard, 1962, 1964). The BLRI is a 64-item self-report questionnaire that is designed to assess the Rogerian therapeutic conditions of empathy, level of regard, congruence and unconditionality of regard provided by the therapist from the client's perspective. The BLRI has been used across many research studies and has been shown to be a valid and reliable measure of the client perceptions of the therapeutic conditions (Simmons, Roberge, & Kendrick, 1995) and has been found to be highly associated with other measures of therapeutic alliance (Salvio, Beutler, Wood & Engle, 1992)), such as the Working Alliance Inventory (Horvath & Greenberg, 1986). Completion of the BLRI is done by rating a 6-point scale that ranges from definitely false, 1, to definitely true, 6. Subscale domains include providers' positive regard ("She/he feels a true liking for me"), empathy ("She/he usually senses or realizes what I am feeling"), and unconditionality ("How much she/he likes or dislikes me is not altered by anything that I tell her/him about myself").

As well as the version mentioned above, the 'other-to-self' (SO) version which assesses the other person's provision of therapeutic conditions to self, there are a number of other versions of the BLRI which assess perceived levels of the

therapeutic conditions from a number of different vantage points. The 'myself-to-other' (MO) version is designed to assess one's own perception of the provision of therapeutic conditions to other. In addition there is also an observer version that is designed to be used by observers observing two people in a relationship (a therapeutic relationship or personal relationship).

Internal reliability of the B-LRI has been found to be high for each of its subscales with average correlations of .84 for the empathy subscale, .91 for regard, .74 for unconditionality and .88 for congruence (Gurman, 1977). Test-retest reliability has also found to be high with .83 for empathy, .84 for regard, .80 for unconditionality of regard and .86 for congruence subscales (Gurman, 1977). The BL-RI has also been found to predict therapeutic outcome (e.g. Watson and Geller, 2005).

2.2.8 Measures Summary

It would seem that many of the above inventories have much in common in that they tap into the collaboration between client and therapist. Others measure the relationship in a different way by being designed to assess the facilitative conditions of the person-centred approach. Therefore, these inventories, questionnaires and scales are not all assessing the same thing, as each one emphasizes, or puts more weight on, one aspect or another. The CALPAS and VTAS are designed to assess negative elements of therapy whereas supportive and commitment elements are captured by the Penn scales. It would appear that agreement, collaboration and consensus are more accurately measured by the WAI. Moreover, as some measures are underpinned by specific theoretical conceptualizations of the alliance such as the Penn, WAI and CALPAS, other scales are very different as they are designed to

assess alliance constructs (VPPS and VTAS). If the above differences between different measures were not enough, the number of items differs considerably across all the measures from 10 items on the Penn Helping Alliance Scales to 80 items on the VPPS. Each measure also has a different number of subscales attempting to assess different numbers of alliance dimensions. The Penn Scales have two dimensions, the WAI three and the CALPAS, four. The only core aspect that all these measures have in common is that they, to some degree or another, are designed to assess relational variables between the therapist and client – which variables they assess largely depends on the measure in question. In spite of this, there are high associations between the many different alliance and therapeutic relationship measures indicating that many relational variables, although conceptually different, may well actually assess very similar aspects of the client therapist union. Of course one of the biggest issues is whether an inventory is actually measuring what it is designed to.

2.3 What is the relationship between relational depth and outcomes?

2.3.1 Introduction

All the relational depth studies mentioned above provide considerable evidence that relational depth exists. However, the above studies are limited in that firstly, although qualitative studies may well provide rich data to evidence the existence of relational depth, they do not address the association between relational depth and outcome in the same way as a quantitative methodology would. In addition, whilst a quantitative methodology does not always provide us with the rich information that qualitative does, it can more easily provide evidence from a much larger sample. Therefore, to be able to investigate more objectively whether relational depth is

related to outcome, one needs to look at quantitative evidence as well as qualitative evidence.

In light of the above, this section focuses on the quantitative studies that exist that provide evidence that relational depth is related to outcome. Therefore the studies reviewed in this section are those quantitative-based studies that focus on a direct measure of relational depth and outcome as well as those that indirectly assess relational depth and outcome. This is because at the time of writing, relational depth-outcome studies are rare. Indeed, at the time of writing, there is only one study, that of Leung (2009), that directly assesses experiences of perceived relational depth and outcome.

Therefore as well as reviewing direct and indirect measures of relational depth and outcome, I also review those studies that assess the therapeutic relationship and outcome. As the therapeutic relationship is seen to be essentially consisting of the therapist's provision of the person-centred facilitative conditions of empathic understanding, congruence and unconditional positive regard, it can be seen to be closely related to the concept of relational depth. I include in this section therapeutic relationship-outcome studies.

This section is divided into two; firstly outcome studies that focus on direct measures and indirect measures of relational depth and, secondly those studies that focus on other client-therapist relationship variables namely the therapeutic relationship and the alliance. In the first, outcome studies that focus on direct measures of relational

depth are those that specifically are designed to assess Mearns and Cooper's (2005) definition of relational depth and how it relates to outcome. Studies which look at indirect measures of relational depth are those which assess the mutual experiencing of the core conditions and outcome. In the second, studies focusing on other client-therapist variables reviewed including those studies that focus on how outcome is predicted by the therapeutic relationship by the alliance.

2.3.2 Direct and Indirect Measures of Relational Depth and Outcome

2.3.2.1 Direct measures of relational depth and outcome.

Leung (2009) conducted a quantitative-based study that focused on both clients' and therapists' experiences of moments of relational depth. An online questionnaire was developed which enquired whether the individual had ever experienced relational depth, the frequency of relational depth and the perceived importance of relational depth on therapeutic outcomes. To do this Leung (2009) included in the questionnaire a definition of relational depth (as put forward by Mearns and Cooper, 2005). He then included a question that asked participants whether they had experienced relational depth according to this definition. Also included in the questionnaire were questions that asked the participant whether they thought relational depth (as defined by the included definition) was related to outcomes of therapy and how important such moments are for personal change.

In Leung's (2009) study, a total of 168 participants completed the questionnaire with 140 reporting to be therapists and 28 as clients. Seventy-two of the therapists also completed the questionnaire twice when they completed a client version of the questionnaire (drawing on their experiences as a client). Another 19 respondents

also completed the questionnaire a second time drawing on their additional experiences of therapy with another therapist. This meant that there were 259 completed questionnaires. Over 88% of the total responses reported they had experienced relational depth. Ratings of importance of relational depth (on a scale from 1 to 7) for therapeutic outcome, was high (5.73) with little difference between therapists and clients. Similarly, ratings of the extent to which relational depth had a high enduring effect were also high (mean = 5.87) with again, little difference between clients and therapists. Other findings were that therapists more than clients, and females more than males, were more likely to report to having experienced relational depth.

However, this study was limited by the fact that it utilized non-independent data (some participants completed the questionnaire more than once, firstly as a therapist and then as a client) and that it did not use conventional tried and tested outcome measures (such as CORE); the assessment of outcome was conducted by using three questions ('During the process of therapy how important do you think that these moments of "relational depth" are for change? In your own experiences, how important do you think that these moments of relational depth are for the outcomes of therapy? In your own experiences, to what extent do you think that these moments of "relational depth" have an enduring effect?')

Despite limitations, this study is the first quantitative study that has investigated whether clients and therapists experienced relational depth and which also investigated the relation between relational depth and outcome. In light of this,

Leung's (2009) questionnaire can therefore be seen to be a direct measure of relational depth. This study would indicate that it would be beneficial to explore relational depth and outcome further using a quantitative methodology.

In light of the fact that relational depth theoretically involves the mutual experiencing of the core conditions of empathy, unconditional positive regard and congruence, I specifically focus on studies that look at how outcome may be related to the mutual experiencing of such conditions. What makes these studies different is that as well as the therapists being assessed for their provision of these core conditions, clients are also assessed for the provision of such. Therefore what is being assessed is something different to what is generally thought of as the therapeutic relationship, which is only the therapist's provision of such conditions. In this light, such mutuality studies deserve to be reviewed separately to therapeutic relationship-outcome studies.

2.3.2.2 Indirect measure of relational depth and outcome.

The mutual experiencing of Rogers' (1957) therapeutic conditions and the experience of relational depth are comparable concepts. Indeed, relational depth is seen to be a state of being or condition that is caused by the mutual experiencing of high degrees of the therapeutic conditions. A key difference is that relational depth is seen to be a state where the mutual experiencing of the therapeutic conditions are present to a high degree and not just simply the mutual presence of these conditions. In addition, relational depth is viewed as a specific moment or event, as opposed to the general quality of the relationship. In this vein the experience of relational depth is different to that of the mutual experiencing of the therapeutic conditions. However,

on the premise that relational depth is the mutual experiencing of high levels of the therapeutic conditions of empathy, congruence and unconditional positive regard, it can be seen that mutuality is an indirect measure of relational depth.

Murphy (2010) researched the mutual experiencing of Rogers' (1957) conditions of therapeutic change and how these relate to positive therapeutic outcome. In the study, the Barrett-Lennard Relationship Inventory (RI) was used (see below for a more comprehensive description of the RI) to assess the mutual experiencing of the therapeutic conditions. Here both client and therapist versions were administered. The Clinical Outcome Routine Evaluation (CORE) was used to assess outcome.

Murphy (2010) found that compared to therapists' views, clients' views of the quality of the therapeutic relationship, were a stronger predictor of outcome. Other findings in the same study showed a significant interaction between client and therapist views of the quality of the therapeutic relationship. This meant that when both clients and therapists perceive there to be mutually high levels of the therapeutic relationship conditions (empathy, unconditional positive regard and empathy) the association between the client view of the relationship and outcome is stronger; in other words the stronger the conditions (as perceived by the client and therapist), the stronger is the positive outcome.

Murphy's (2010) findings suggest that the perception of mutually high levels of congruence, empathy and unconditional positive regard are able to predict positive therapeutic outcome. Given that relational depth is, theoretically, the mutual

experiencing of the therapeutic conditions, this study can be seen to be indirectly assessing relational depth and outcome (please see below for further discussion as to how the RI can be seen to indirectly assesses relational depth). In light of this, it seems a reasonable assumption that an experience of relational depth may also be associated with positive therapeutic outcome.

2.3.3 Other client-therapist relational variables and outcome.

2.3.3.1 Overview

The most documented client-therapist relational variables include alliance-outcome studies and therapeutic relationship-outcome studies. Whilst some studies do not differentiate between these two concepts and often refer to the therapeutic relationship as included in the alliance, here I distinguish between those studies that focus on the alliance and those that focus on the therapeutic relationship.

In this section I wish to firstly look at outcome studies which assess the therapeutic relationship and how it is related to outcome. As already mentioned, I refer to the therapeutic relationship as that aspect of the client-therapist encounter that includes the therapist's provision of the person-centred facilitative conditions of empathy, unconditional positive regard and congruence. This is in attempt to focus on outcome literature that is most closely linked to that of relational depth and outcome simply because it is these person-centred conditions which underpin the theory behind relational depth. The second part of this subsection reviews literature that focuses on the alliance (the alliance as seen as goals tasks and bonds) and how it relates to outcome.

Despite the alliance being something quite conceptually different to relational depth it is still inextricably linked simply because, like relational depth, it is still a relational variable within the client-therapist relationship. Furthermore, as the evidence suggests that the alliance and the therapeutic relationship are related, the alliance may be seen to be another linked to relational depth. I therefore also include in this section alliance-outcome studies.

2.3.3.2 Therapeutic Relationship and Outcome

Reviews and Meta-analyses

Lambert and Barley (2001). Lambert and Barley (2001) conducted a review that summarized over one hundred studies concerning the therapeutic relationship and its relation to psychotherapy outcome. With regard to client outcome, they focused on four areas; these were extra therapeutic factors (e.g. spontaneous remission, fortuitous events and social support), expectancy effects (e.g. placebo), specific therapy techniques (e.g. biofeedback), and common factors (therapeutic relationship factors (e.g. empathy, warmth and the therapeutic relationship)). On average, the size of contribution that each predictor made to outcome was that 40% was due to outside factors, 15% to expectancy effects, 15% to specific therapy techniques, and 30% of variance was predicted by the therapeutic relationship/common factors. One of their key findings was that among those factors most closely related to therapist activity, it was the common factors that were most significant in contributing to therapy outcome. Lambert and Barley (2001) concluded that emphasizing the relationship along with other common factors is more likely to enhance positive therapeutic outcome than focusing on techniques. However, the authors also add that in regarding client-therapist relationship factors, conceptually differentiating between

therapist variables (such as interpersonal style, and therapist attributes), facilitative conditions (empathy, warmth, congruence) and alliance, is difficult. They add that these concepts are not always mutually exclusive or distinct. It is therefore unclear from this study, exactly which proportion of relationship variables the person-centred facilitative conditions contributed to outcome.

Karver, Handelsman, Fields & Bickman (2006). In a meta-analyses the researchers looked at 49 studies of youth and family therapy. The analysis essentially included studies that looked at individual, family or parent treatments for clients under the age of 18. A total of the 19 studies looked at how the qualities of empathy with warmth or genuineness predicted outcome. Karver et al refer to such qualities as counselor interpersonal skills. They found that the relation between counselor interpersonal skills and youth outcome varied with effect sizes from .06 to 1.32 but the weighted mean effect size was .35. Karver et al point out that this is consistent with comparable adult studies that look at therapist empathy and treatment outcome, namely Bohart, Elliott, Greenberg and Watson (2002).

Elliott, Bohart, Watson and Greenberg (2011). In a meta-analysis which focused on the relations between therapeutic outcome and empathy, Elliott, Bohart, Watson and Greenberg (2011) included published research where empathy measures were used and where empathy was related to a measure of therapy outcome. The meta-analysis included 224 separate tests of empathy-outcome association with 59 different samples and a total of over 3, 500 clients. Two sets of analysis were conducted with one by effects and one by studies. The effects analysis used the 224 separate tests

and the studies analysis used the 57 separate studies. Pearson correlation were used to measure effect sizes. The study's results suggested that empathy accounts for an effect size of .30 at the study level and .22 at the analyses level. The authors concluded that empathy accounts for about 9% of the variance in therapy.

Farber & Doolin (2011). Farber & Doolin conducted a meta-analytic review of positive regard and outcome psychotherapy studies. The inclusion criteria for the meta-analysis was that studies had to a) identify positive regard as unconditional regard, positive regard, warmth, nonpossessive warmth, affirmation or acceptance b) Positive regard was considered as a predictor of outcome c) included quantitative information regarding an effect size d) clients were adults or adolescents e) clients were individuals (not couples or groups). As a result 18 studies were included from a total of 44 articles. Effect sizes were computed for each of the 18 studies. The aggregate effect size was .27 and the authors state that this suggests that positive regard has a moderate association with psychotherapy outcome.

For the core condition of congruence / genuineness, Kolden, Klein, Wang and Austin (2011) undertook a meta-analytic review of 16 studies. Inclusion criteria of the studies were that the study had to include quantitative information in order to calculate effect size. Estimates of the effect sizes in all of the 16 studies ranged from -.26 to.69. The weighted aggregate effect size for congruence was .24 which is a medium effect size.

Single Studies

Watson and Geller (2005). In a study which specifically looked at Rogers' facilitative conditions by using the Barrett-Lennard Relationship Inventory (BLRI), Watson and Geller (2005), wanted to investigate the association of clients' ratings of the relationship conditions (as assessed by the BLRI) with outcome and working alliance. The Working Alliance Inventory (WAI) was used to assess working alliance. More specifically they also wanted to explore any mediator effects that the working alliance had on the relation between outcome and the relationship conditions. Participants were 66 clients who were suffering from depression. Clients either had Cognitive Behaviour Therapy (CBT) or Process-Experiential Therapy (PET). The BLRI predicted outcome on 4 different outcome measures (where pre-therapy scores on each were controlled for), the Beck Depression Inventory (BDI), Inventory of Interpersonal Problems (IIP), Rosenberg Self-Esteem Inventory (RSE) and the Dysfunctional Attitudes Scale (DAS). The WAI also predicted outcome, in the same way, but on 3 of the 4 outcome measures (not the IIP). The BLRI scores were a significant predictor of WAI scores. However, when the WAI was controlled for, the BLRI was no longer a significant predictor of outcome on any of the measures. This would imply that the impact of the relationship conditions (on outcome) was mediated by the alliance on 3 out of 4 of the outcome measures; in other words when working alliance was controlled for the relationship conditions were no longer a significant predictor of outcome. Unexpectedly no significant differences were found between PET and CBT therapists for the BLRI subscales of unconditional acceptance or congruence. Empathy approached significance where PET therapists scored slightly higher than their CBT counterparts. Watson and Geller (2005) discuss that the mediating role

that the alliance plays with the relationship conditions and outcome 'indicates that the relationship conditions facilitate the development and maintenance of a good working alliance across different therapies with different theoretical assumptions' (p.30).

Zuroff and Blatt (2006). Also using the BLRI, Zuroff and Blatt (2006) conducted a study that used data from the National Institute of Mental Health Treatment for Depression Collaborative Research Program (TDCRP). Participants were 250 clients suffering from depression. Researchers utilised the BLRI as a measure of the relationship conditions. Outcome measures were the BDI, the Hopkins Symptom Checklist (HSC), the Global Adjustment Scale (GAS) and the Social Adjustment Scale (SAC).

Using multilevel modeling they found that clients with high BLRI scores showed a quicker decrease in maladjustment than those with low scores indicating that a positive therapeutic alliance is strongly predictive of rapid decline in maladjustment. In addition, clients and therapists' BLRI scores predicted a decrease in maladjustment at 18 months follow-up indicating that a positive therapeutic relationship early in treatment also predicted greater improvement after therapy and that the therapeutic relationship has long-term effects.

Zuroff and Blatt (2006) also controlled for a wide range of client characteristics (age, gender, marital status, severity of maladjustment, length of episode, depression severity and personality disorder characteristics). None of these characteristics were

found to eliminate the effect of the therapeutic relationship. However, the personality disorder characteristic, perfectionism, was significantly correlated with residual BLRI scores (scores were residualised to control for early reduction in maladjustment). The authors state that this would suggest that perfectionism may lead to an impaired therapeutic relationship and worse long-term outcome. However, the most important finding of this study was that a positive therapeutic relationship early in therapy contributes directly to positive therapeutic outcome regardless of the therapy approach.

2.3.3.3 Alliance and outcome

Of all the different aspects of therapy, relationship variables have been the most researched and documented in the literature with regard to predicting outcome. As relational depth is seen to be a relationship variable, it is highly appropriate that alliance-outcome research is looked at in detail. This is especially the case since research into relational depth is still in its infancy and there is currently (at the time of writing) no research that focuses on relational depth and outcome that uses conventional outcome measures (such as CORE). However, among the different relationship variables, the therapeutic alliance (aka working alliance or helping alliance) has been the most documented and has been found to contribute the largest portion to client outcome in psychotherapy.

Since the time of Sigmund Freud, there has been a consistent interest, amongst the psychotherapy community, in the alliance between client and therapist. Horvath and Bedi (2002) state that the reason for this interest is the wealth of evidence in existence that suggest that many different psychotherapies produce similar beneficial

effects. They also suggest that the surge of interest is due to the impact of Roger's person-centred theory, which focused on the therapeutic relationship between client and therapist. Martin, Garske and Davis (2000) offer a reason for research interest in the alliance by suggesting that it is the inability of researchers to find consistent differences across orientations in therapeutic outcome. Furthermore, it would seem that there is strong evidence that suggests that most orientations of psychotherapy are effective. Stiles, Shapiro and Elliott (1986), for example, concluded that "differently labeled therapies have demonstrably different behavioural contents, yet appear to have equivalent outcomes" (p.175).

In 1913, Sigmund Freud was perhaps one of the first psychotherapists to study how the client-therapist alliance could enable successful therapy (Freud, 1958b). He believed that positive transference towards the therapist helped foster positive outcome and that successful psychoanalysis could not happen '…until an effective transference has been established in the patient, a proper *rapport* with him. It remains the first aim of treatment to attach him to it and to the person of the doctor…' (p.139).

However, it has mainly been in the last thirty years where alliance-outcome research has truly become far more established. Such a wealth of outcome research has largely come about due to the numerous types of inventories, measures and scales in existence that are designed to assess client-therapist relationship variables (see below for a comprehensive view of such). Whilst some measures are a global assessment of the therapeutic alliance, others assess component parts of it. As there is such a

large number of studies which focus on alliance and outcome, there is not the space here to review them all. As a result, only the most relevant studies and meta-analytic studies are reviewed. I will look at alliance-outcome studies as well as other relationship-outcome studies, especially those which are most related to relational depth such as those studies which focus on the relationship between Roger's therapeutic conditions and outcome.

By the early 1990s there had already been a wealth of research that focused on the relationship between the alliance (either 'working alliance', therapeutic alliance' or 'helping alliance') and therapeutic outcome. In turn, because of this wealth of research many researchers focused on meta-analyses of alliance-outcome. A key figure in alliance-outcome research is Adam Horvath and colleagues.

Horvath and Symonds (1991). In a meta-analysis of 24 studies, which focused on working alliance and outcome (Horvath and Symonds, 1991) found that, on average, about 6.5% of the variance in outcome was due to alliance factors. Here, this outcome is generally measured in terms of symptom improvement and client satisfaction within treatment. They also found that the clients' quality assessments of working alliance, were most predictive of treatment outcomes than therapists' assessment and observers' assessment of such. Furthermore, type of therapy, length of treatment, whether the research was published or number of participants in the study did not influence the association between alliance and outcome.

Horvath and Luborsky (1993). In another review, Horvath and Luborsky (1993) focused on how the alliance was characterised over time, what predisposes people to

develop a strong alliance, as well as factors that develop a positive alliance. Their findings were that the strength of the alliance remains relatively stable over time and can also gain strength over time but that this gain in alliance strength does not appear to be related to outcome. The researchers also concluded that intrapersonal and interpersonal client variables (such as difficulty maintaining social relationships, poor family relationships) are often associated with poor alliance.

Luborsky (1994). In a review which looked at factors which explain the predictive success of the alliance, Lester Luborsky (1994) concluded that 1) a positive rather than a negative alliance was associated with positive outcome, 2) regardless of whoever makes the rating, therapeutic alliance is related to outcome, 3) the client's rating of the alliance predicts outcome better than the therapist's, but that 4) some studies reported a lack of relationship between perspectives (observers, clients and therapists), 5) type of therapy is not predictive, and 6) the time course of therapy is not predictive of outcome.

Martin, Garske and Davis (2000). In a later review, Martin, Garske and Davis (2000) conducted a meta-analysis based on 79 studies. Their findings seem to suggest that the overall relationship between therapeutic alliance and outcome was moderate, having an effect size of .22. Their findings also imply that there was no difference between raters (client, therapist or observer) to predict outcome. Similarly the association between alliance and outcome did not appear to be influenced by the type of outcome measure, type of outcome rater, time of alliance assessment, type of alliance rater, type of therapy or whether the study was published.

Horvath and Bedi (2002). Horvath and Bedi (2002) presented a summary of research which included the data used from Horvath and Symonds (1991) and Martin, Garske and Davis (2000) meta-analyses as well as an additional 10 studies. They found that across all studies the median effect size was .25. Time of alliance assessment was found to have some influence on outcome; early alliance assessment was found to be a marginally better predictor of outcome than mid-therapy assessment but that late alliance assessment is higher than mid-therapy. Horvath and Bedi (2002) put this down to contamination with therapy benefits. The relation between alliance and outcome did not appear to be influenced by type of measure, professional training of therapist or type of therapy.

Horvath, Del Re, Fluckiger and Symonds (2011). More recently Horvath, Del Re, Fluckiger, and Symonds (2011) conducted a considerable research synthesis which included over 200 research reports based on 190 independent data sources. The overall correlation between alliance and treatment outcome was .275. Results were comparable to previous studies and indicated that the impact of the alliance on therapy outcome is 'ubiquitous irrespective of how the alliance is measured, from whose perspective it is evaluated, when it is assessed, the way the outcome is evaluated, and the type of therapy involved' (p.13).

2.4 Chapter Summary

From the above it would appear that there is strong evidence to suggest that relational variables in therapy are related to outcome. However, it is not always clear as to whether relational variables refer to working alliance, the therapeutic relationship or a combination of both. Also, there appears to be some inconsistency

between studies with not all finding evidence that relational variables and outcome are related. Many of these issues are often thought to be due to the many different ways in which the client-therapist relationship is actually assessed. There is strong qualitative evidence that relational depth exists and some qualitative evidence that it may lead to better outcome. There is some quantitative evidence that relational depth is related to outcome but measures were not vigorously tested for validity and reliability evidence. However, there is quantitative research evidence that both the therapeutic relationship in general and alliance in particular are related to positive therapeutic outcome. There also appears to be numerous measures and inventories designed to assess client-therapist relationship variables but none with evidence of validity and reliability that assess relational depth. Chapter 4 focuses on the various ways such relational variables are assessed and thus looks at the various alliance measures in existence as well as other more radical approaches to assessing the client-therapist relationship. Before looking at other relational variables, however, the next section (Chapter 3) focuses on the main philosophy of this research project in the context of epistemology and major epistemological approaches.

Chapter 3: Review of Epistemology in Context of this Research Project

3.1 Introduction to Chapter

The object of this research project is to gain knowledge about a specific area of psychotherapy. The research method we use to acquire knowledge depends on our own view of the world and our own view about what knowledge actually is and how we possess knowledge. Epistemology is essentially the study of how we know what we know. Therefore, epistemology is very relevant to research and has been defined as

> …the study of knowledge and justified belief. As the study of knowledge, epistemology is concerned with the following questions: What are the necessary and sufficient conditions of knowledge? What are its sources? What is its structure, and what are its limits? As the study of justified belief, epistemology aims to answer questions such as: How we are to understand the concept of justification? What makes justified beliefs justified? Is justification internal or external to one's own mind? Understood more broadly, epistemology is about issues having to do with the creation and dissemination of knowledge in particular areas of inquiry (Steup, 2011).

This section reviews the relevant literature regarding epistemology and in so doing looks at the various philosophical approaches to knowledge. In addition, it reviews and outlines the philosophical approach to the current project and therefore puts the epistemology behind this particular research project in context with other approaches to epistemology. In particular, it looks at different philosophies or movements

focusing on how we gain knowledge. Such philosophies include Empiricism, Rationalism, Positivism and Postpositivism, which are essentially terms that describe groups or systems of philosophical ideas. Some of these share characteristics with others and some are quite distinct. However, they all have something in common in that they each have something to say about 'how we know what we know'.

3.2 Rationalism

Rationalism posits that real knowledge is *a priori* (Martin, 2010). This means that knowledge can be gained by applying intuition, innate knowledge or reason without necessarily referring to particular evidence or experience (Markie, 2012). Rationalism has its roots in Platonian philosophy (Martin, 2010) and as a result has perhaps a longer history than most other philosophies. Therefore, rationalists assume that there is an objective reality without there being solid evidence of it (this is also an empiricist stance). Moreover, rationalism holds the view that this reality is such that knowledge can be obtained about it through intuition or innate knowledge (Markie, 2012). In this vein, rationalists therefore also hold the view that conclusions can be drawn about reality through logical argument and reason.

Three central tenets are central to rationalist philosophy. The Intuition/Deduction Thesis, the Innate Knowledge Thesis and the Innate Concept Thesis (Markie, 2012). The first of these, the Intuition/Deduction Thesis posits that it is possible to have knowledge by intuition alone (Markie, 2012). It sees intuition as a form of rational insight in that knowledge is gained by logical deduction and takes the view that the process of deduction helps us make conclusions from intuited premises through logical arguments. The second, the Innate Knowledge Thesis, takes the premise that

it is possible to have knowledge in a particular area simply because we are rational by nature (Markie, 2012). It assumes we can gain knowledge *a priori* (before, despite or independently) of experience. This is in opposition to the idea that we can gain knowledge *a posteriori* (after experience). The difference between the Intuition/Deduction thesis and the Innate Knowledge thesis concerns how this a priori knowledge is acquired. In the former it is gained through intuition and deductive reasoning and in the latter through our innate nature. The Innate Concept Thesis is really no different to the Innate Knowledge thesis in that it sees that some of what we know is not gained by experience but by our rational nature.

In light of the above, the main characteristics of rationalism is that a priori knowledge is in some way superior to a posteriori knowledge (Cardinal, Hayward and Jones, 2004).

3.3 Empiricism

Empiricism is different from Rationalism even though there are some similarities (which I outline below). Empiricism, in it most extreme form holds the view that we can only really know something through experience (Cardinal, Hayward and Jones, 2004). Where for rationalists reason and intuition is the source of knowledge, for empiricists, it is experience that is seen as the source of knowledge (Cardinal, Hayward and Jones, 2004). However, sensory experience is fallible and subjective and there is a degree of doubt (scepticism) even within some branches of empiricism. The English philosopher John Locke was more moderate in his approach. Although an empiricist, he believed that knowledge which is observable could be acquired both

externally and internally (in the mind). His view contrasted with rationalism in that Locke believed the mind begins as an empty slate (Locke, 1976) .

3.4 Positivism

Central to positivism is the view that science is the way to gain knowledge and to be able to understand the world in order that it can be predicted and controlled (Trochim, 2001). It was August Comte, a French philosopher, who was a pioneer of positivism and coined the phrase (Bourdeau, 2011). Comte saw that a variety of different philosophies improved and contributed to our understanding over the course of history. He pioneered the use of the scientific method to the 'soft' sciences of sociology and psychology. A staunch empiricist, Comte did not believe in the existence of any metaphysical knowledge and was also very critical of religion and theology.

Later, Hans Reichenbach (1891-1953) who was essentially an empiricist, developed a philosophy that was rooted in scientific methodology (Glymour and Eberhardt, 2011). He staunchly believed in the scientific results that scientific methodology brought about. In doing so he was also committed to objectivity and realism. With regard to epistemology, Reichenbach believed that science over-emphasised technical problems (Glymour & Eberhardt, 2011). He believed that a way to rectify this was for philosophers to closely analyse physics in order that they are able to explain scientific theories in a philosophical way (Murzi, 2001).

Around the same time Ludwig Wittgenstein (1889-1951) put forward philosophical ideas which essentially applied modern logic to metaphysics and in doing so brought

a new understanding in the relations between the world, thought and language (Biletski and Matar, 2011). In his most famous work, *Tractatus Logico-Philosphicus* in 1921, Wittgenstein presented a 'solution' to what he called the problems of philosophy (Biletski and Matar, 2011). He made seven basic propositions: 1) The world is everything that is the case, 2) What is the case, the fact, is the existence of atomic facts, 3) The logical picture of the facts is the thought, 4) The thought is the significant proposition, 5) Propositions are truth-functions of elementary propositions. (An elementary proposition is a truth function of itself). 6) The general form of truth-function is $[p, \xi, N(\xi)]$. This is the general form of proposition. 7) Whereof one cannot speak, thereof one must be silent (Biletski and Matar, 2011).

3.5 Postpositivism

The main philosophy underlying the methodology used in the research process here is postpositivism. As the name suggests, this philosophy came after positivist philosophy. Positivism emphasized provable facts and directly observable phenomena (Trochim, 2001). Positivist philosophy also subscribed to the 'scientific method' and drew heavily on quantitative methodology and experiments. Postpositivism, on the other hand, rejects the central tenets of positivism but only to a degree. Where positivism embraced realism and would posit that scientists were capable of an objective truth, postpositivism recognizes that the main aims of science are concerned with accuracy but also recognizes that these aims can never be achieved perfectly. This particular view is referred to as critical realism and is still a part of postpositivism. Trochim (2001) states that postpositivism emphasizes the importance of multiple measures and observations in order to address the fact that all measurement is fallible. Ryan (2006) also states that "investigating your own

epistemologies and understanding how they affect you as a researcher is an essential part of the post-positivist approach" (p.18).

3.6 Chapter Summary

The brief outlines of the main epistemological approaches places the current research project in perspective. The research philosophy for this current project is not rationalist because it does not rely on intuition or innate knowledge, it is not empiricist because it does not rely solely on experience and it is not positivist because it does not rely solely on the scientific method. Quantitative methods are used extensively in this research project and whilst it is recognized that such methods are usually underpinned by positivist principles, they are appropriate to the research aims in this project. I therefore seek to follow a postpostivist approach by using multiple measures and observations in order to investigate the concept of relational depth. For example, from a postpositivist perspective, this study acknowledges that the research conducted here is capable of being fallible. In addition, it also acknowledges there may be variables that are not directly observable. Consequently, various statistical methods and approaches will be used which facilitate the idea that some variables are latent and therefore not directly observable. For example, exploratory Factor Analyses and the Rasch Model are used extensively in this research project.

Chapter 4: Review of Methodology in the Context of this Research Project

4.1 Introduction to Chapter

The aims of this chapter are to document in detail all the various methodologies used throughout this research project. Because the main objective of this research project is to find evidence of validity for the Relational Depth Inventory (RDI), a large part of this chapter focuses on, and documents, various psychometric methods used when testing an inventory for reliability and validity. It first describes and documents the various theories and concepts as to what reliability and validity are. It does this by looking at the many different theories of validity by giving a potted history of what has constituted reliability and validity over the last 50 years or so. It then looks at data collection methods which focus heavily on online data collection due to the fact that a large part of the research conducted for this project utilized a website for data collection. The final part of this chapter looks at significant events research mainly because the research conducted here falls under the general category of significant event research; because this current research focuses on a moment or event it thus falls into significant event research. This particular kind of research possesses a unique research method and so this is more relevant to research methodology than more general research literature.

4.2 Reliability and validity

4.2.1 Introduction to Reliability and Validity

When creating a psychological instrument, such the Relational Depth Inventory, it is imperative that research is done to try to ensure it is as reliable and valid as is reasonably possible. Insofar as validity is concerned, if one is to create questions or

items assessing relational depth then one has to also create items that assess the components parts of relational depth. Insofar as reliability is concerned, measures are usually taken to ensure that the test and its items are consistent. Acceptable levels of both reliability and validity are needed if a test, measure, inventory or scale is fit for purpose. Furthermore, reliability and validity are different things and one does not equate to the other. A scale is not valid if it is not reliable but reliability does not equate to validity. Evidence of reliability is necessary for validity but not sufficient.

There are many ways to test for reliability and some are straightforward such as the test-retest approach. More complex methods include statistical procedures of internal consistency. Whichever method is used, the testing of reliability is usually less complex and more objective than that of validity. The concept of validity, or the appropriateness of how a test score is interpreted, is far more complex and is continuously evolving (Gregory, 2007). It is also the most fundamental consideration in developing and evaluating tests (American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education NCME, 1999). Despite being thought of as the most important aspect of a test, the concept of validity has changed over time and has notoriously been a subject of discussion (Gregory, 2007).

In this section I intend to provide a substantial additional review of the literature, to include a major, in-depth section on reliability theory, validity theory and their practice. I incorporate multiple sources from relevant authors, including the various

editions of the Standards for Educational and Psychological Testing (Standards from here on) such as the latest version (AERA, APA, NCME, 1999).

4.2.2 What is Reliability?

Reliability of a test has been described as 'the proportion of variance attributable to the true score of the latent variable' (DeVellis, 2003, p.27). One way of looking at reliability is that if one were to administer a test to the same sample of people twice, how consistent (similar) would the test results be? In this instance, the extent of the similarity is the extent of the reliability of the scale. The extent to which a test can generate consistent scores is a testament to the test's precision. Therefore, reliability is essentially concerned with how similar two things are. In this subsection I look at the concept of reliability by beginning with the history of the concept and then go on to demonstrate various methods of testing for reliability.

4.2.3 History and Theory of Reliability

Karl Pearson (1901) devised a statistical calculation that intended to calculate the strength of association between two random variables and is more commonly known as the Pearson Product Moment Correlation. A colleague of Pearson, at University College London, Charles Spearman, saw the concept of reliability as how similar two concepts were. In 1904 he published *The Proof and Measurement of Association between Two Things* (Spearman, 1904). In his article, Spearman (1904) set about addressing the lack of suitable methods for calculating correlation when using non-parametric data. This calculation is commonly known as Spearman's Rho. His work built on that of Pearson in that it added another method of calculating the strength of

association between two variables. Most notably, he formulated a new statistical procedure that of the non-parametric correlation coefficient.

The work of Spearman (1904) came to the attention of Edward Thorndike in the United States, who published *An Introduction to the Theory of Mental and Social Measurements* (Thorndike, 1904, cited in Thorndike, 1916). This was a comprehensive publication that documented in detail his theories of the reliability and validity of scales. Since the work of Spearman (1904) and Thorndike (1904, cited in Thorndike, 1916) other statisticians and developers have put forward methods and analyses. Perhaps one of the most important of these are Kuder and Richardson (1937) who introduced, in one publication, a number of new reliability coefficients. One of the most notable is the KR20 which is a statistical procedure designed to test the internal reliability of a test which has dichotomous data. Later, Lee Cronbach (1951) and Lee Cronbach and colleagues (Cronbach and Meehl, 1955) continued to develop reliability methods and focused identifying error in measurement, most notable, Cronbach's alpha, a statistical procedure designed to test the internal consistency of tests which had polytomous data. The theory of reliability carries on developing and more recently researchers have focused on developing models that address items response patterns by using Item Response Theory. These statistical procedures are discussed and outlined below.

4.2.4 Models of reliability

This subsection focuses on the different models of reliability. I focus on the three main approaches to test reliability; the test-retest method, parallel form method and the internal consistency method.

4.2.4.1 Test-retest method

The test-retest method of reliability is used to check the reliability of an instrument over time. It specifically is designed to identify any margin of error that might occur as a result of taking the test at a different time. Test-retest reliability is fairly easy to put into practice as one just simply administers it on two separate occasions to the same sample of people. Then the two sets of scores would be compared by using a variety of methods. All methods of correlation analysis compare pairs of observations and in most cases compare pairs of scores. In test-retest analysis a pair of scores are analysed for each person. A common method would be to conduct a correlational analysis to produce a correlation coefficient. One of the most common is the Pearson Product Moment Correlation Calculation (Coolican, 1999). One of the limitations of this method is concerning the type of measure being tested. For example if the test or measure is designed to test things like mood or feelings, one might expect a change in scores over time due to the person feeling differently from one time to the next. Another problem is carryover effects. These occur when the experience of completing the test first time round affects the completion second time round. For example, participants may get used to the test in question or they may complete the test the second time with less naivety (and know the answers). Therefore this may not always be the best method.

4.2.4.2 Parallel forms method

A parallel form method involves the use of two inventories that are designed to measure the same thing. The reason for this method is to address some of the problems that test-retest methods suffer from. In other words the researcher might find themselves in a position in which they wish to test participants more than once

but also want to eliminate carryover effect. A typical use of parallel forms therefore would be in something like a spelling test where the tester would want to present different sets of words with each set having the same difficulty. This method is not practiced very often simply because most researchers do not have two versions of the same test. In fact it may not be possible to have two versions of the same test simply because there may be only one way to test or measure a certain construct. When there is only one version of a measure in existence, testing for reliability can be done using methods of internal consistency.

4.2.4.3. Internal Consistency

According to DeVellis (2003) central to internal consistency is the homogeneity level of test items, which is established by calculating the strengths of association between test items. Internal consistency of a test refers to the intercorrelations among items within the same test (Kaplan & Sacuzzo, 2009). There are a number of ways of testing for internal consistency.

Split-half method. The split-half method tests the reliability of a test or inventory by, as the name suggests, splitting it in half and having one half completed and then the other half. The results from the two halves are then compared. However, the problems with this are that different halves may have differing difficulties. In this case one half would have odd numbered items and the other half the evens. In calculating the reliability of the tests various analyses can be used including the Pearson Correlation or the Spearman-Brown Prophecy Formula.

Pearson Product Moment Correlation. The Pearson correlation coefficient can be defined as the covariance which exists between two variables divided by the product of their standard deviation (Coolican, 1999). A coefficient can be between -1 and 1 with the former indicating a perfect negative correlation and the latter a perfect positive correlation. In reality however, such perfect correlations are extremely rare especially in the social sciences.

Spearman-Brown Prophecy Formula. The Spearman-Brown Prophecy Formula (Brown, 1910, Spearman, 1910) is often used when using the split-half method above. That is when there are two sets of items, where each set of items is less than (usually half of) what the usual test would be. Using a Pearson Correlation would not be a sufficient estimate of reliability simply because there are half the number of tests scores and reliability increases the more test items there are. In other words, a Pearson Correlation would be an underestimate of reliability. The Spearman-Brown Prophecy Formula allows for the calculation as if the whole test had been taken (rather than half).

Coefficient Alpha. DeVellis (2003) defines alpha as 'the proportion of a scale's total variance that is attributable to a common source, presumably the true score of a latent variable underlying the items.' (p. 31). The are two main types of statistic that each give a coefficient alpha, the Kuder-Richardson 20 Formula (KR_{20}) and Cronbach's coefficient alpha. The KR_{20} or the Kuder-Richardson 20 Formula is named so because it was the 20^{th} formula that Kuder and Richardson presented in their publication on reliability (Kuder and Richardson 1937). It is used for testing a set of

items that are dichotomous, where there is, for example, scores of 0 or 1 (0 indicating non endorsement and 1 indicating endorsement).

While the KR20 formula is used for dichotomous data, Cronbach's (1951) coefficient alpha is used for polytomous data. (Polytomous data is data which can belong to several levels – it is non-binary, in other words.) It is probably one of the most widely used methods of testing for internal consistency of a test scale.

4.2.4.4. Interrater reliability

When one needs to know the exact agreement between raters, correlation coefficients are not very helpful (Howitt and Cramer, 2008). An interrater reliability index needs to provide an index of the overlap of ratings and should be sensitive to the problem of raters who do not varying their ratings (Howitt and Cramer, 2008). Unlike the raw data that are used in many of the methods described above, the raw data from interrater scores are usually two or more sets of scores from two or more raters and are thus not questionnaire or test scores. In most cases, interrater scores are derived from the observations of raters where each rater has observed a participant's behaviour, or description of feelings. There are numerous ways in which any given set of interrater scores can be tested for reliability. One of the most well-used is something called the Kappa coefficient or Cohen's Kappa (Cohen, 1960). This is a measure of agreement between a pair of raters and takes into account of the amount of agreement that would be expected by chance alone. The expected numbers agreeing are found as in chi-squared tests, by row total times column total divided by grand total (Howitt and Cramer, 2008). The Spearman-Brown Prophecy Formula can also be used to calculate interrater reliability for pooled or averaged ratings

(Brown, 1910 and Spearman, 1910).

4.2.5 What is validity?

Central to reliability is the extent to which a variable can influence a set of items whereas central to validity is whether the variable actually causes the influence (or covariation) in the items (DeVellis, 2003, p.49). According to Messick (1990) validity

> is an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the *adequacy* and appropriateness of interpretations and actions based on test scores or other modes of assessment. The principles of validity apply not just to interpretive and action inferences derived from test scores as ordinarily conceived, but also to inferences based on any means of observing or documenting consistent behaviors or attributes. (p.1)

Others have described validity as concerning 'whether the variable is the underlying cause of covariation' (DeVellis, 2003, p.49). In other words, if the items on a reliable measure are assessing what they are designed to assess, any change or differences in scores can be attributed to change or differences in the construct that the test is designed to measure. Therefore if a test respondent scores highly on a valid anxiety scale one week, and then scores low the following week, this variation in score would be due to the variation in anxiety in the test respondent. Testing a measure in order to find evidence that it is valid is therefore of prime importance as it can provide the test developer with information regarding the extent to which a test is assessing what it is supposed to measure. However, the definition of validity and

underlying theories have not always been similar to those described above. The history and theory of validity has changed over time. The following section focuses on how the history of test developers have defined validity over the last 60 years.

4.2.6 History and Theory of Validity

During the early 1950s the widely accepted view of validity was that the extent to which a test was valid was the extent to which it correlated with another criterion. In this sense, a test was seen as having many criteria or many validities (Gullikson, 1950). However, the mid 1950s through to the sixties saw a considerable evolution of the validity concept. In a key publication, Cronbach and Meehl (1955) put forward a view of validity and categorized four specific types of validation: content validity, concurrent validity, predictive validity and construct validity. Content validity referred to the content of the test itself such as questionnaire items and tasks and whether they were representative of the concept they were designed to assess (Cronbach and Meehl, 1955). Concurrent and predictive validity were both subtypes of criterion-related validity and referred to the extent in which a test correlated with an external criterion. Such an external criterion could be completed at the same time as the test thus assessing concurrent validity or at some time in the future thus assessing predictive validity. The fourth type of validity was construct validity and was the most elusive of the four. (It is important to mention at this point that although there were four types of validation, this was seen to be a tripartite system because concurrent and predictive validities were subtypes of criterion-related validity.) Cronbach and Meehl (1955) stated that construct validation takes place whenever interpretations of a test score are seen to be a measure of a construct that cannot be "operationally defined" (such constructs are today conceptualised as latent

variables). Construct validity, was (and still is) seen to be the extent to which there is evidence that interpretations of test scores are appropriate (Bond and Fox, 2007).

Cronbach and Meehl (1955), in the same paper also put forward the idea of the nomological network which is a system of test development designed to develop validation methods. The term nomological means lawful (and is derived from Greek *nomos* and *logos*) and so the nomological network means the lawful network. Essentially what Cronbach and Meehl (1955) were trying to do was to link concept and theory with observations.

Later, Campbell and Fiske (1959) extended the tripartite view and put forward the multitrait-multimethod approach. Here the authors addressed the need for many types of validity evidence and introduced two other types of validity; convergent and discriminant validity. In practice, this approach assessed two or more traits or psychological constructs using two or more methods to assess each. This meant that each trait would be assessed by more than one method and each method would assess more than one trait. This approach therefore put each method under scrutiny and was designed to uncover any possible weaknesses in the assessment of any given psychological construct. This was seen to be a very robust method during its time and still is today in that it is able to test methods as well as the constructs in question. However, this method is rarely put into practice largely because of the resources and time needed.

In 1966 the Standards for Educational and Psychological Testing (*Test Standards* form here on) was first published jointly by the APA, AERA, and NCME (Goodwin and Leech, 2003). This replaced two earlier publications, namely Technical Recommendations for Psychological Tests and Diagnostic Techniques by the APA in 1954 as well as Technical Recommendations for Achievement Tests by a committee representing AERA and the National Council on Measurement Used in Education (NCMUE, NCME), published by the National Education Association in 1955 (Goodwin and Leech, 2003). Here, the meaning of validity changed in that the focus came away from correlations with other tests, and validity was now defined as the degree to which a test or measure yielded information that could be used for a particular purpose (Goodwin and Leech, 2003). However, the 1950s trinity view of validity was still included in the publication and remained the predominant view of validity over the next two decades.

It was not until the 1980s that significant challenges were made to the concept of validity. Lee Cronbach (1980, cited in Kvale 1995), rather than discussing the validity of a test, put forward the idea of the validity of interpretation and that such validity cannot be simply established by 'a research monograph or detailed manual (p.107).' Another notable psychometrician during this era was Samuel Messick who put forward a more radical approach. Messick (1980) was critical of the tripartite view of validity, stating that the implication that validities come in different types only leads to confusion and over-simplification. In the same article he further indicated that many test-users make the mistake of focusing on one or another of the 'types' of validity rather than on the inferences from test scores (Messick, 1980).

His argument was that there was really only one type of validity, namely construct validity and that content and criterion validity were not types of validity but types of evidence (Messick, 1980).

In 1985 the *Test Standards* were updated and still included the tripartite view of validity as well as using category labels but warned test-users about such categories:

> …the use of category labels [for types of validity] does not imply that there are distinct types of validity or that a specific validation strategy is best for each specific inference or test use. Rigorous distinctions between the categories are not possible. Evidence identified usually with the criterion-related or content related-related categories, for example, is relevant also to the construct-related category. (AERA, APA, & NCME, 1985, p. 9).

This was a radical move forward in the concept of validity in that the emphasis was now moving away from the idea that there were many validities and moved toward the idea that there was only one type of validity that of construct validity. Around the same time Cronbach (1988) stated that the 30-year old idea of three types of validity is an idea whose time has gone.

By the 1990s the idea of different types of validity was simply not acceptable to many psychometricians and many contributed to the new unified concept of validity. Messick (1995) stated that the traditional concept of validity (as dividing into three types) failed to take into account the value implications of score meaning as a basis for action and of the consequences of score use. In the same publication, Messick

(1995) further described six aspects of construct validity, namely content aspects, substantive aspects, structural aspects, generalizability, external aspects and consequential aspects. Jeri Benson (1998) offered a framework for conducting a strong program of construct validation. She was also keen to point out that content and criterion-related validity "should not be considered separate and distinct forms of validity" (p.11). She advised the test-user that weak programs are characterised by a heavy dose of semi-related correlations between the construct in question and other constructs which appear to be 'raked' together. She further continued that strong programs, on the other hand, are characterized by three stages; substantive (e.g. generate theoretical definitions), structural (e.g. item/subscale definitions) and external stages (e.g. relationship with other tests).

The new approaches to validity during the 1990s inspired changes to the Test Standards. During 1993, the three sponsoring organizations of the Standards set about revising and updating the 1985 version and by 1999 a new version was published. In looking at the 1999 version in comparison to the 1985 version, it is noticeable that the latest one is around double the size and the structure has changed somewhat. The 1999 Test Standards contains three parts: Part I Test Construction, Evaluation, and Documentation, Part II Fairness in Testing and Part III Testing Applications. The 1985 Standards had 16 chapters organized in four parts. Looking closely at the chapters in both books the chapters seem very much the same just organized differently. However there appear to be some significant changes. One such change includes a new chapter on fairness (Chapter 7) that essentially addresses that tests have a lack of bias and that there is equitable treatment in the testing

process. Another change is the merging of two chapters into one; there is now one chapter on employment testing and licensure and certification (Chapter 14). Two other chapters that have also now merged are the chapters on clinical testing and test use in counselling. These two subjects are now one chapter (Chapter 12). This was done in order to include all types of psychological testing and present them in one coherent chapter rather than separate ones.

The changes mentioned above are small but significant changes. However the most important changes include those surrounding validity. In the 1999 Test Standards, a definition of validity was decided upon which encapsulated the view of many psychometricians.

> Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests…The process of validation involves accumulating evidence to provide a sound scientific basis for proposed score interpretations. It is the interpretations of test scores required by proposed uses that are evaluated, not the test itself' (p.9).

> Validity is a unitary concept. It is the degree to which all the accumulated evidence supports the intended interpretation of test scores for the proposed purpose. (AERA, APA, & NCME, 1999, p. 11).

Since this most recent version of the Test Standards, many test developers and psychometricians have put forward ways in which the test developer might interpret

and put the new Standards into practice. Yun and Ulrich (2002), for example, rather than referring to different validities, refer to different types of evidence; content-related evidence, criterion-related evidence and construct-related evidence. With content-related evidence they state this should be derived from operational definitions and panels of judges (experts in the field of the construct in question). Regarding criterion-related evidence, they refer to this as predictive and concurrent related evidence with measures put in place where there are predictive criteria and concurrent criteria. With regard to construct-related evidence they state there are three stages; defining the construct, formulating a hypothesis and using appropriate statistics.

Cook and Beckman (2006) come away from any 'content' 'criterion' labels and further embrace validity as a unitary concept. They state there are five sources of evidence that support construct validity; content (construct definition), response process, internal structure, relations to other variables and consequences. The content concerns the actual items or tasks of a test. With regard to response process, this refers to examining the actions and thought processes of test-takers, which should go some way to uncover whether the intended construct is being assessed. Internal structure refers to reliability where statistical methods such as Cronbach's alpha are used. Relations to other variables simply refers to the extent to which the measure in question relates to other measures. The consequences element largely refers to examining intended or unintended consequences (e.g. where participants may not understand a questionnaire item).

Outside of psychometrics, many other test-users follow aforementioned validity theory especially the conceptualization as outlined by the various editions of the Standards. For example, Rowe and Mahar (2006) present a paradigm for the process of construct validation in exercise science where they suggest that accumulation of evidence is on three levels: a definitional stage, a confirmatory stage and a theory-testing stage. The definitional stage largely involves defining the construct and the confirmatory stage is essentially gathering criterion-related evidence as already mentioned above. They suggest the theory-testing stage can involve a combination of using the nomological network and structural equation modeling (SEM).

The steps used in this project are intended to reflect the fact that validity is widely seen as a unitary concept. There will be no investigation of different 'types' of validity but rather investigation of different types of evidence. Every step taken therefore attempts to develop an inventory that has accompanying evidence, which test that items are being interpreted as appropriately as possible.

4.2.6 Models of Validity

This subsection focuses on how the validity of a test is assessed. As already mentioned, in testing the RDI for validity, I see validity as a unitary concept and therefore see there being only one type of validity that of construct validity.

However there are generally seen to be different kinds of validity evidence. This is essentially what this subsection focuses on.

4.2.6.1 Face-Value Related Evidence for Validity

As already mentioned, the latest Test Standards (1999) sees validity as a unitary concept and that there is only one type, which is construct validity. However, still in use is the term *face validity*. It is technically not a form of validity, but I wish to address the aspects of a test that address how valid the test appears. In light of the latest Test Standards (1999), I wish to refrain from using the term face validity and instead propose a new term that addresses the validity evidence that concerns the appearance of a test; *face-value related evidence*. I intend for this term to refer to the face-value of the items of the test which simply means the extent to which the items look valid. Face-value related evidence is particularly relevant to this project due to the method in which test items were originally created (by using participants' descriptions of relational depth which were subjected to grounded theory analysis). In this vein, the initial RDI items could be seen to have a considerable amount face-value related evidence.

As the name suggests, face-value related evidence can be seen as the extent to which test items appear to be measuring what they are supposed to. In other words face-value related evidence refers to the extent that a test item or set of test items look as though they are testing what they are designed to assess. This is quite important as Kaplan and Saccuzzo (2009) point out that appearances can help motivate those completing the test or inventory because they will see that it is relevant. In addition, it is of prime importance that items are written in a way that participants understand and are comfortable with. Wording is of utmost importance and must not confuse or be ambiguous. Consequently it is important to have clear guidelines as to how items should be worded. Frey, Peterson, Edwards, Pedrotti and Peyton (2005) wanted to

create a list of experimentally tested rules for item-writing. To do this they examined the quality of classroom assessment textbooks. Four different types of validity concerns emerged: 1) potentially confusing wording or ambiguous requirements, 2) the problem of guessing, 3) test-taking efficiency and 4) controlling for testwiseness. Frey et al (2005) then identified list of 40 item-writing rules (these were the 40 that had most consensus). These are given below.

1. "All of the Above" should not be an answer option.
2. "None of the Above" should not be an answer option
3. All answer options should be plausible.
4. Order of answer options should be logical or vary.
5. Items should cover important concepts and objectives.
6. Negative wording should not be used.
7. Answer options should include only one correct answer.
8. Answer options should all be grammatically consistent with stem.
9. Specific determiners (e.g. *always*, *never*) should not be used.
10. Answer options should be homogenous.
11. Stems must be unambiguous and clearly state the problem.
12. Correct answer options should not be the longest answer option.
13. Answer options should not be longer than the stem.
14. Items should use appropriate vocabulary.
15. In fill-in-the-blank items, a single blank should be used, at the end.
16. Items should be independent of each other.
17. In matching, there should be more answer options than stems
18. All parts of an item or exercise should appear on the same page
19. True-false items should have simple structure.
20. True-false items should be entirely true or entirely false.
21. There should be 3 to 5 answer options.
22. Answer options should not have repetitive wording.
23. Point value of items should be presented.
24. Stems and examples should not be directly from textbook.
25. Matching item directions should include basis for match.
26. Answer options should be logically independent of one another.
27. Directions should be included.
28. Questions using the same format should be together.
29. Vague frequency terms (e.g. *often*, *usually*) should not be used.
30. Multiple-choice stems should be complete sentences.
31. There should be an equal number of true and false statements.
32. True-false statements should be of equal length.
33. Individual items should be short.
34. In matching, answer options should be available more than once.
35. Number of answer options should be < 7 for elementary age tests.
36. Number of answer options should be <17 for secondary age tests.
37. Complex item formats ("a and b, but not c") should not be used.
38. All items should be numbered.
39. Test copies should be clear, readable and not hand-written.
40. Stems should be on the left, and answer options on the right.

Most of the rules above are relevant to the current project with the exception of Item 7 ('Answer options should include only one correct answer') due to there being no 'correct' answers on the RDI. DeVellis (2003) also comments about test items stating that very lengthy items should be avoided, reading difficulty should be considered, double-barreled items should be avoided and ambiguous pronoun references should also be avoided.

4.2.7.2 Content-Related Evidence for Validity

Content-related evidence simply refers to the content of the test in question or more specifically the content of the items. Content-related evidence can be thought of as sampling adequacy or the extent to which the set of items are representative of the particular content domain the test intends to cover (DeVellis (2003). In this way, it is not so different from face-value related evidence because it focuses on the meaning of the test and test items.

4.2.6.3 Criterion-Related Evidence for Validity

Criterion-related evidence refers to the extent to which items on a test have an empirical association with some criterion. The criterion in question may be another similar questionnaire to the one that is being developed. However, there must be strong evidence that the criterion has strong validity. The criterion in this way is the standard that the newly developed test must literally measure up to if it is to be considered as valid. There are two types of criterion-related evidence; predictive-related evidence and concurrent-related evidence.

Predictive-Related Evidence.

Predictive-related evidence refers to a type of criterion-related evidence where the test in question has would be tested for its predictive quality in relation to the criterion. In this case the measure being tested would be given first and the criterion given subsequently Kaplan and Sacuzzo (2009) refer to this predictive quality as a 'forecasting function' (p. 138). In the psychotherapy literature some alliance inventories have predictive-related evidence in that their scores are predictive of outcome questionnaire scores. The outcome questionnaires would be the criterion in this case.

Concurrent-Related Evidence.

Another type of criterion-related evidence is concurrent-related evidence which refers to evidence gleaned from some criterion that is given concurrently with the measure in question. For example, if one wanted to obtain criterion-related evidence for the Working Alliance Inventory (WAI) one might give another measure of the alliance at the same time the WAI is given.

Validity Coefficient

A validity coefficient is a coefficient that has been produced through conducting a correlation calculation between a criterion and the test in question. Therefore, the validity coefficient is a direct indication of the extent to which there is criterion-related evidence for validity.

4.2.6.4 Construct-Related Evidence

Messick, (1995) described construct validity as 'the evidential basis for score interpretation' (p.7). In one sense, all types of validity evidence are construct-related evidence. However, this does refer to a particular type of validity evidence but all

the types of validity evidence already mentioned can contribute to construct-related evidence. What this means is that to obtain such evidence one must use a number of methods; no one activity or test or analysis can establish construct-validity evidence. Test developers have been very clear with this kind of evidence stating that it 'must be investigated whenever no criterion or universe of content is accepted as entirely adequate to define the quality to be measured.' (Cronbach and Meehl, 1955, p.282). Construct-related evidence can bee seen as the accumulation of various types of evidence that makes explicit what the test in question means.

Convergent-related Evidence

Convergent-related evidence refers to evidence that would show that the test in question is strongly associated with another test measuring the same or similar thing. In this case, evidence would be obtained that showed the two measures converging or 'narrowing in' on the same thing (Kaplan and Sacuzzo, 2009).

Discriminant-related Evidence

Discriminant-related evidence is really the same thing as convergent-related evidence but in reverse. If a test has discriminant (sometimes called divergent) evidence for validity it will not be associated with something that it shouldn't be associated with. For example, if a test is designed to assess depression, it should have at least a very low, zero or negative association with a scale designed to measure happiness.

Multitrait-Multimethod Matrix

Multitrait-Multimethod Matrix (MMTM) is really more of a method of providing construct-related evidence than a type of evidence in itself. However, it can be loosely described as a type of evidence as it combines many of the different

evidences already mentioned above and therefore can be seen to come under the general heading of construct validity evidence. In order to be able to provide evidence that a psychological test has construct-related evidence for validity, convergence and discrimination have to be evidenced. Designed by Campbell and Fiske (1959), the method involves collecting discriminant and convergent evidence and using different methods to acquire each of these types of evidence. This would mean assessing whether traits that are supposed to be unrelated, are unrelated in reality and, similarly, those that are expected to be related, are related in reality. The MMTM is essentially a matrix of correlation coefficients.

Reliability Diagonals (monotrait–monomethod). Unlike a conventional correlation matrix, rather than 1s running through the diagonal, the estimate of the reliability of each measure is added to the diagonal cells (these are indicated by bracketed text and located across the diagonal from top left to bottom right). The figures that would be entered in these cells would be the alphas of internal consistency (such as coefficient alpha). In other words these are measures of internal reliability and one of a number of methods can be used depending on the method in questions. For example, for observations one might consider interrater reliability measures using the Cohen's Kappa correlation (as outlined above in the in the section on reliability). The values should be considered as monotrait-monomethod correlations.

Validity Diagonals (monotrait-heteromethod). The bold text indicates validity diagonals – that is those cells which contain correlations between the same traits but where different methods have been used. One can see there is one validity diagonal

for each method. If the measure (Me) has strong evidence for validity then one would expect high correlations between the same trait whatever the method used.

Heterotrait-monomethod triangles. What the figures have in common in these triangles is method; they do not have trait in common and are indicated by darkly shaded groups of three figures. Often, correlations are high when using the same method even though different traits are being measured. This is sometimes due to a strong methods factor.

Heterotrait-heteromethod triangles. The lightly shaded sets of three cells (triangles) represent heterotrait-heteromethod triangles. These figures are in lightly shaded cells and represent associations between measures which share neither trait nor method. Consequently, we would expect figures here to be the lowest of the whole matrix.

Monomethod blocks. As the name suggests, figures that are in the cells in these blocks have one method. There should be the same number of blocks as there are methods of assessment.

Heteromethod block. The figures in these blocks represent those associations between measures that do not share the same method.

To summarise, according to the Campbell and Fiske (1959), if there is strong evidence that the measure in question (the one being tested for validity) has validity qualities then the following criteria should apply.

1. In the validity diagonals figures should be high (or at least significantly higher from zero).
2. The values in the validity diagonal should be higher than the neighboring values in the column and row in the hetero-trait and hetero-method triangles.
3. A trait should correlate higher if a different method is used to measure the same trait than if the same method is used to measure a different trait; if two same traits have been measured differently those two traits should correlate more highly with each other than two different traits which have been measured in the same way.
4. In all of the heterotrait triangles of both the monomethod and heteromethod blocks, the same pattern of figures should exist. For example, in the matrix above, in the first column of each block the value decreases as one moves down the column. Furthermore, the values also decrease across the first row from left to right.

It can be seen then that using the MTMM can enable both convergent and discriminant validity evidence to be assessed. However, the model is extremely complex and to be able to test a measure using different traits and method is costly and extremely time consuming and consequrntly rarely carried out.

Threats to Construct validity

Messick (1995) states that there are two major threats to construct validity: construct 'underrepresentation' and 'construct-irrelevant variance.' In underrepresentation, the test in question is too narrow in that it (for example) it does not assess the construct fully enough. Construct-irrelevant variance is when the test is too broad in that it

may have items that assess a different but related construct to the one being assessed. Such items would evidence reliability but are not necessarily assessing the intended construct.

An example of 'underrepresentation' might be when a test designed to assess depression only has items that assess mild depression and lacks items that assess a deep, clinical depression.

With regard to construct-irrelevant variance Messick (1995) puts forward that there are two forms: 'construct-irrelevant difficulty' and 'construct-irrelevant easiness.' The former is when aspects of the test, such as particular items, are outside of the construct in question and cause difficulty to some test-takers. An example of this may be where test-takers do not understand questions or may be uncomfortable at endorsing some test items. Messick (1995) adds that

> In general, construct-irrelevant difficulty leads to construct scores that are invalidly low for those individuals adversely affected (e.g., knowledge scores of poor readers). Indeed, construct-irrelevant difficulty for individuals and groups is a major source of bias in test scoring and interpretation as well as of unfairness in test use. (p. 10)

With regard to construct-irrelevant easiness, Messick (1995) states that this occurs when extraneous clues in test items leads test-takers to endorse an item but in ways that may not be relevant to the intended construct. This may happen when test-takers

are familiar with the subject of the question (and perhaps the wording of the question) when compared with other test items. Construct-related easiness can therefore cause participants to have scores that are invalidly high. To summarise, construct-irrelevant difficulty elicits validly lower scores for some test takers while construct-irrelevant easiness leads to invalidly higher scores for other test takers. It is paramount therefore that test developers are constantly examining the test details, wording of test items, administration methods and scoring techniques in an effort to reduce the impact of sources of invalidity.

4.2.7 Classical Theories in Testing for Reliability and Validity

4.2.7.1 Classical Test Theory

Classical test theory was first named as such in a paper published by Melvin R. Novick (Novick, 1966) in which he outlined the basic assumptions of classical test theory. Classical test theory aims to aid our understanding and improve the reliability of psychological tests. Underlying the concept of classical test theory is the assumption that each person in a given test has a *true score*. Further, the score that a person gives on any test is an *observed score* and may not be totally representative of a true score. That is because such a score may have an element of *error* caused by a number of potential factors such as an imperfect test, or a participant that guesses a response. Therefore it is assumed that *observed score* = *true score* plus some *error*. In this vein, classical test theory can be seen to be virtually the same thing as *true score theory*.

The word "classical" distinguishes it from other more contemporary theories and approaches (such as item response theory which sometimes includes the term

"modern" as in "modern latent trait theory") and refers to the fact that it was one of the first approaches to psychological testing.

Classical test theory consists of a whole range of mainly statistical procedures and analyses. Each analysis or procedure is designed in such a way that it has margin of error built in to it. For example, the simple T-test compares two sets of data (typically two sets of scores). The rationale behind a T-test is that the difference between means must somehow be compared with the variation within the groups (Coolican, 1999). The question as to whether the two sets of scores are significantly different from each other rests on the calculation of the standard error of the mean difference. The Standard Error is calculated by finding the mean of the two means then calculating the standard deviation of that mean. This standard deviation of the mean is called the standard error (SE) or standard error of the mean (SEM).

Calculating the standard deviation or the standard error has little use unless used in the calculation of *confidence intervals*. A confidence interval can be seen as a margin of error. Confidence intervals are obtained by multiplying the standard deviation or the standard error by a ***t*** value obtained from a table of *t* values. A confidence interval is in fact a range of values within which a given percentage (e.g. 95%) of observations could be expected to lie. Put another way, if a comparison was made many times with different pairs of samples of people (different pairs of data sets) the confidence interval allows us to infer that in 95% of cases the mean could be expected to fall within the range of values given by the confidence interval. The *t*-test is typical of the kind of analyses that one would undertake within the classical

test approach in that incorporated into the statistical procedure is an element which theoretically allows for a margin (or confidence interval) of error.

4.2.7.2 Generalizability Theory

Generaliazability theory was developed by Lee Cronbach and his colleagues (Cronbach, Rajaratnman and Gleser, 1963) and was different to classical test theory in that instead of seeing error as coming from one source, it saw error as having many different sources. According to DeVellis (2003) Generalizability theory (G theory) 'provides a framework for examining the extent to which one can assume equivalence of a measurement process across one or more dimensions.' (p. 46). Similarly, Kaplan and Sacuzzo (2009) describe generalizability as referring to 'the evidence that the finding obtained in one situation can be generalized – that is applied to other situations.' (p. 146). To put these definitions in perspective, a very relevant generalizability question regarding the current study concerns whether the data collected from the online participants is generalizable to the data collected from the research clinic participants. In the current study then, a dimension is mode of administration in which there are two modes of administration; via website and via research clinic. Within G theory dimensions are seen as *facets* and are potential sources of variation (DeVellis, 2003). Further, facets are seen to have different levels which in the current study consisted of website administration and research clinic administration. In G-theory, a test score is seen to be a sample from a *universe of admissible observations.* Such a universe is seen to consist of all possible observations on what is termed as the *object of measurement* (typically a test respondent).

In order to investigate whether one can generalize across samples, and thus treat data collected from different samples as one set of data, one can conduct a Generalizability-study or G-study. Such a study in generalizability terms, would aim to establish the extent to which scores are generalizable across different levels of a facet. This is often carried out by using ANOVA.

4.2.7.3 Measurement Theory

Measurement theory comes under the general umbrella term of applied mathematics. Central to the concept of measurement theory is that the actual measurement being taken is not necessarily the actual thing that is being measured. Therefore if we want to be able to assess a particular construct, for example, the relationship between the construct and the measure must be taken into account. Stevens (1946) introduced a theory of levels of measurement and described nominal, ordinal, interval, and ratio scales of measurement. The reason behind making such classifications was so that it would help guide statistical procedures. Such levels of measurement are briefly outlined in Table 4.1.

Stevens (1946, 1968) was passionate about using appropriate statistics with data obtained from measurement and was critical if he thought statistical procedures were inappropriate to the raw data. He believed that inferences from measurement and statistical procedures should represent the truth.

> To the British committee, then, we may venture to suggest by way of conclusion that the most liberal and useful definition of measurement is, as one of its members

> advised, "the assignment of numerals to things so as to represent facts and conventions about them." (Stevens, 1946, p. 680)

What Stevens (1946) was getting at was that inferences drawn from measurement data that had been analysed using statistical methods must represent the truth. An example where an inference from measurement may be misleading concerns temperature; 50 degrees Celsius is not twice as hot as 25 degrees Celsius even though numerically 50 is twice 25. This was the point Steven (1946) was making; that statistical procedures must be appropriate to the data that is collected. Put another way, the conclusions drawn from any statistical analysis should not be reliant upon how we may have, for example, assigned scores to a Likert scale because there are many different ways to assign scores to such a scale. The inferences that are made from a statistical analysis should say something about the truth, not about any (subjective) decision as to, for example, how Likert scales were assigned scores.

Table 4.1 Levels of measurement

Scale (determination)	Brief Explanation	Examples of Permissible Statistic or Transformation
Nominal (equality)	Numbers or labels are assigned to one or more cases if they are equivalent (e.g. all participants who had Cognitive Behaviour Therapy).	Number of cases Mode Correlation
Ordinal (greater or less)	Numbers are assigned reflecting order (e.g. 95% = 1st, 80% =2nd, 60%=3rd).	Median Percentiles
Interval (equality of intervals)	Numbers are assigned to levels of a variable where such numbers directly reflect the equal increments of the variable (e.g. length - 1=20ft, 2=40ft, 3=60ft).	Mean Standard Deviation Pearson Correlation T-test
Log-interval (equality of log-intervals)	Numbers are assigned to cases that reflect ratios (e.g. miles per hour).	Power transformation
Ratio (equality of ratios)	Numbers are assigned to reflect the difference between ratios	Any linear transformation (unit of measurement is arbitrary).
Absolute (actual number represents measurement e.g. no of participants)	Numbers are assigned which directly reflect the actual measurement.	Identity transformation

Note. Table adapted from Stevens (1946).

4.2.8 Non-classical Approaches in Testing for Reliability and Validity.

4.2.8.1 Item Response Theory (IRT)

The purpose of using a questionnaire to assess a psychological construct is to provide a relatively simple and convenient way to acquire psychological data (Gregory, 2007). If using a classical test theory approach, the overall score of a completed questionnaire is usually calculated by summing each participant's score on a series of questions, where each question has an ordinal rating scale. However, the validity of such scores has been called into question by those proponents of item response theory, including proponents of the Rasch approach to validation testing (Bond and Fox, 2007). The issue is that the classical approach to rating scales assumes all item stems (e.g. item statement or questions) are equally weighted. It also assumes that there are equal intervals between each items anchor point. Item Response Theory, including the Rasch approach, on the other hand, assumes items vary in difficulty and it does not assume equal intervals between rating scale anchor points (Bond and Fox, 2007).

Generally thought as an alternative approach to Classical Test Theory (CTT), Item Response Theory (IRT) is not just one approach but can be seen to be a family of models or mathematical techniques rather than one particular theory (Bond and Fox, 2007). IRT is also known as *latent trait theory* in areas other than educational and psychological testing but often the two terms are interchangeable because there is so much overlap between them (Uebersax, 2000).

It would appear that central to all IRT approaches is the attempt to model an individual's response to an item. Therefore, fundamental to IRT is the notion that

persons are positioned on a latent trait and that this position, along with particular item parameters, determines a person's response to a particular item (Gregory, 2007). This means that the relationship between a person's position on a latent trait and any particular item that the person encounters is a probabilistic one. In turn this also means that item difficulty is central to all IRT models but this is pretty much where the similarity ends as there are many IRT models.

4.2.8.2 One, Two and Three Parameter IRT Models

The following subsection explains in detail three different IRT models; the one, two and three parameter logistic models. It explains in detail how each these of bear resemblance to each other and how they differ. Equations and diagrams are given to aid the reader understand the differences between each of these models.

Item difficulty refers to the strength of the construct being measured 'that is associated with a transition from "failing" to "passing" the item' (De Vellis, 2003, p.139). The one-parameter logistic model (1-PL model) is designed to address item difficulty and is so named simply because item difficulty refers to the single parameter in question (De Ayala, 2009). From item difficulty this model predicts a person's response based solely on how difficult the said item is to endorse. Each item has a different difficulty so the probability of a person with a particular ability of endorsing an item is only determined by the item difficulty. The equation (4.1) of the one-parameter model is expressed below.

$$P(\theta) = \frac{1}{1 + e^{-(\theta - b)}}$$

Equation 4.1 The 1PL Model (Adapted from Yan, 2009 and Gregory, 2007)

The *e* is the base value of the natural logarithm that has the constant value of approximately 2.71828. Therefore, the above equation can be interpreted as the probability of a respondent (*P)*, with trait level (*θ),* of responding to an item, is equal to 1 divided by 1 plus e to the power of the negative of *θ* less difficulty. Mathematically the 1-PL model and Rasch Models are equivalent (Gregory, 2007, De Ayala, 2009, Sick, 2008a). However, the Rasch Model takes a different approach philosophically (De Ayala, 2009). Where the Rasch Model is used to construct the variable of interest, the 1-PL model focuses on fitting the data to the model (De Ayala, 2009). The Rasch Model will be outlined later in the next subsection.

In the two-parameter logistic model (2-PL model) as well as there being the parameter for item difficulty, there is a second parameter for item discrimination. In the 2-Pl model information is calculated as to how well an item discriminates among persons who are at different points logistically. Item discrimination is usually represented by *a.* The equation for the 2-pl model is shown below (Equation 4.2).

$$\mathrm{P}(\theta) = \frac{1}{1 + e^{-a(\theta - b)}}$$

Equation 4.2 The 2PL Model (adapted from Yan, 2009 and De Ayala, 2009)

As can be seen, the equation for the 2-PL is different from the 1-PL model insofar as the negative power to which *e* is raised, is now multiplied by discrimination *a.*

The three-parameter logistic model is simply an extension of the 2-Pl model. This model simply adds another parameter - the *pseudo-chance parameter* (more commonly known as the guessing parameter) (De Ayala, 2009). The symbol *c*

represents the pseudo chance parameter in the equation below. The 3PL equation is expressed below (Equation 4.3)

$$P(\theta) = c + \frac{1-c}{1+e^{-a(\theta-b)}}$$

Equation 4.3 The 3PL Model (adapted from Yan, 2009 and De Ayala, 2009)

The 3-PL Model changes from the 2-PL Model in that *c* (guessing) is subtracted from the numerator after being added to the fraction. When specific values are entered in to the equation and are plotted, an Item Characteristic Curve (ICC) can be plotted. This can help to understand the relationship or interaction between item difficulty and person strength of attribute. Figures 4.1 to 4.3 shows three such curves with each one representing an item with a different difficulty level. Therefore this ICC represents the 1-PL Model.

The darkest line in Figure 4.1 represents the item that is easiest to endorse in that somebody with an average strength of attribute (0 on the horizontal axis) has a 50% chance (0.5 on the vertical axis) of endorsing this item. If persons are going to have a 50% chance of endorsing the other, more difficult, items then they will need to display more of the attribute. For example, if the item represented by the lightest line is to have a 50% chance of being endorsed, the person endorsing it must display a more than average amount of the strength of attribute (a measure 1 on the horizontal axis).

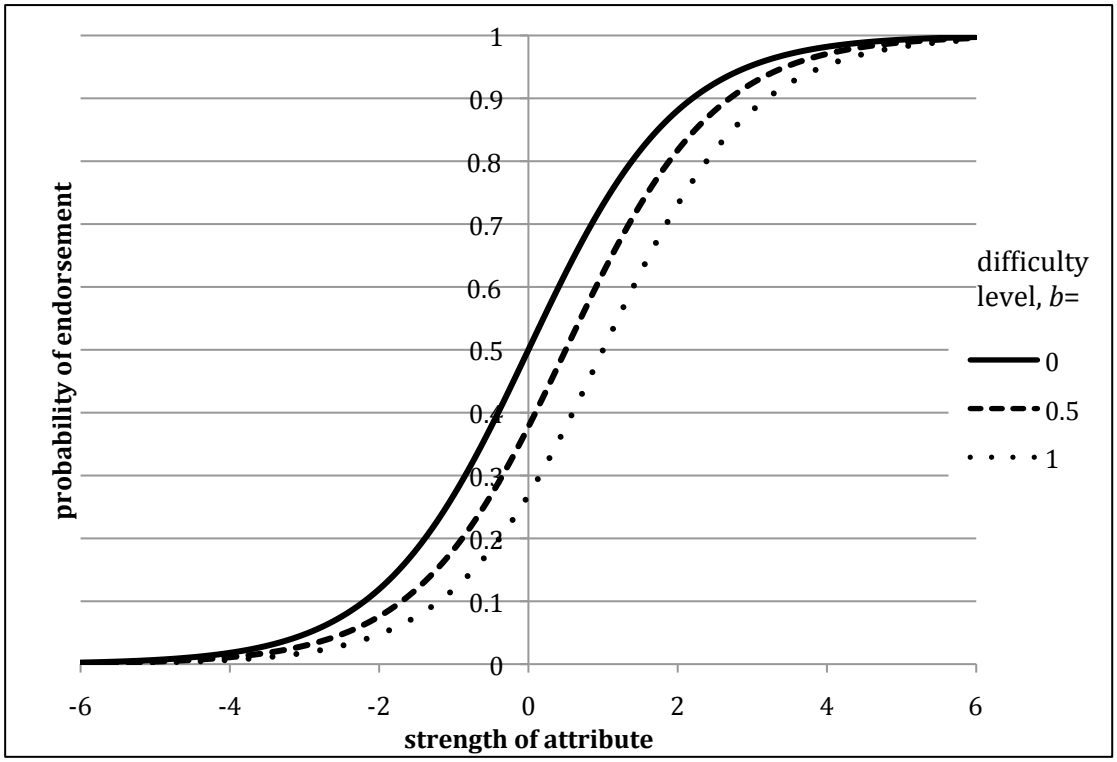

Figure 4.1. Item characteristic curves for the 1-PL Model

For the 2-PL Model, this is represented in Figure 4.2. As can be seen each curve has a different shaped slope which is in contrast to the previous model where each curve had the same slope. Curves that are steeper represent the items that are more discriminating (Yan, 2009). Therefore in Figure 4.2, the darkest line represents the least discriminating item and the lightest line represents the item that discriminates the most.

In Figure 4.3 an ICC represents the 3-Pl model. The three lines represent three items with different guessing parameters. At the -6 point on the horizontal axis the three lines are at different points on the probability of response vertical axis. This means

that persons with little or no strength of the attribute (-6) have differing probabilities (0, 1 or 2) of endorsing an item by guessing.

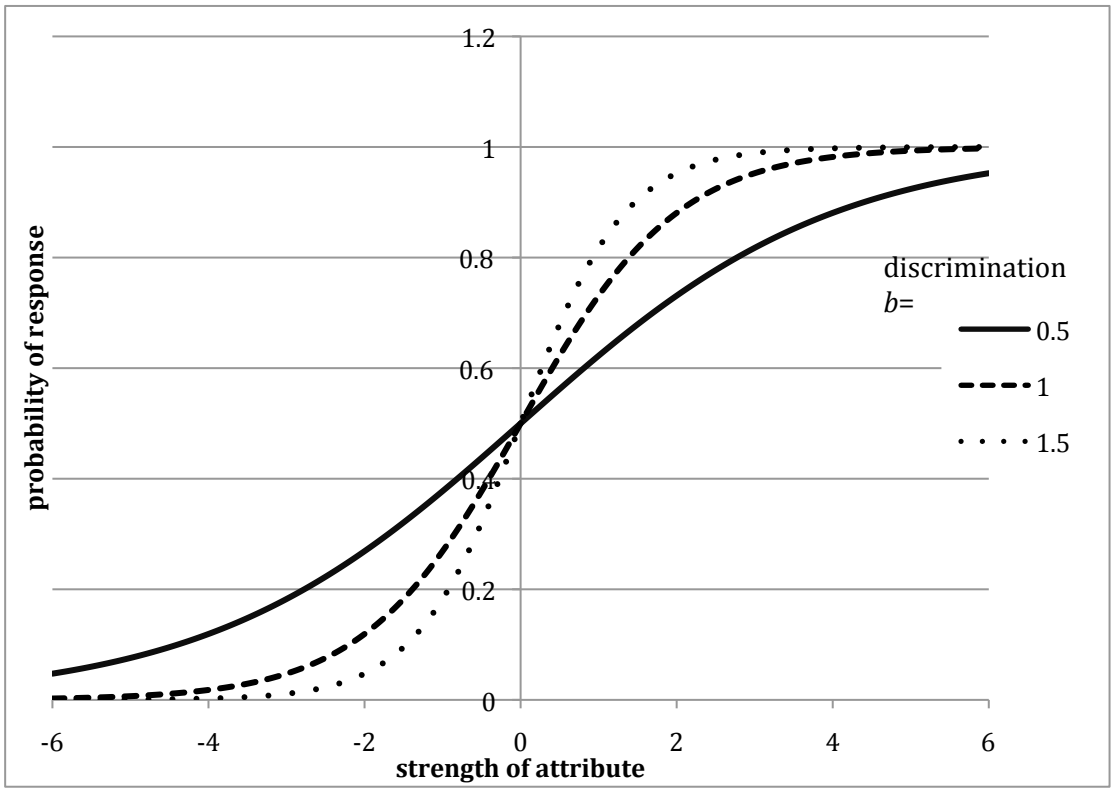

Figure 4.2. Item characteristic curve for the 2-PL Model.

In sum, the one, two and three IRT models have one thing in common in that they all have a parameter for item-difficulty. Further parameters are added to allow for item discrimination and guessing. Another aspect they all have in common is that, unlike Rasch, none of the one, two and three parameter IRT models have a parameter for person agreeableness (often termed person ability). The Rasch model therefore, is arguably an entirely different approach to IRT and may be regarded as not belonging to the IRT model at all. The next part of this section explains the Rasch Model and what Rasch analysis is citing prominent literature regarding the Rasch Model.

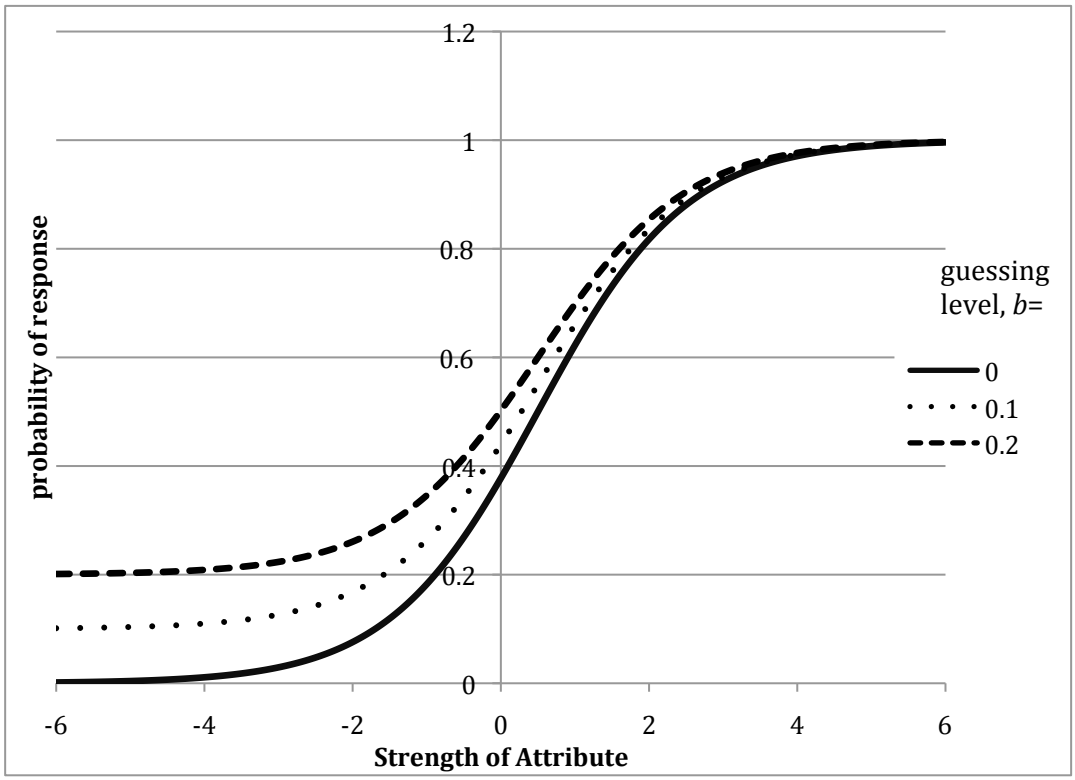

Figure 4.3 Item Characteristic Curves for 3-PL Model.

4.2.9.3 The Rasch Model

Overview of the Rasch Model

The Rasch model is named after its inventor, George Rasch, who was a Danish mathematician (Olsen, 2003). The Rasch approach was originally developed for educational purposes which is why the terms *person ability* and *item difficulty* are often used by Rasch proponents and cited in other Rasch literature (e.g. Bond & Fox). However, the Rasch model can be applied to social sciences or any other discipline where any psychological construct is being assessed (Andrich, 1988, Bond and Fox, 2007)). The Rasch Model builds objective rules of measurement for the human sciences in order to parallel the kind of objective, fundamental measurement that is executed successfully in the physical sciences (Bond & Fox, 2007).

Consequently, it is possible for the human and social sciences to provide information that is more objective (Bond & Fox, 2007).

The Rasch Model approach views each person as having a strength of attribute or ability parameter which is representative of the amount of trait (being measured) a test-taker possesses with each item having a difficulty parameter (Bond and Fox, 2007). Each of these parameters are part of the same latent trait and literally on the same metric. Because of this the Rasch Model can be thought of as a one-parameter model, even though, technically, it has two parameters – item difficulty and person agreeableness (often termed as person ability). Therefore, the answer to an item by any given person is determined by a combination of item difficulty and person ability which together form one parameter.

Mathematically, the Rasch Model is similar to other IRT models and is often thought of as a particular kind of 1-PL model (De Ayala, 2009) However, it is distinct from other IRT models in a number of ways and it is arguable as to whether the Rasch Model actually belongs under the general umbrella of IRT models at all. The ways in which Rasch is different from other IRT models comes down to how it treats person ability, item discrimination and guessing (De Ayala, 2009; Bond and Fox, 2007). Unlike the other IRT models, the Rasch model does not set parameters for guessing or discrimination (De Ayala, 2009). This is because proponents of Rasch see guessing and discrimination as sources of problems that may have an impact on the measurement (Bond and Fox, 2007). Wright (1997) points out that differences in item discrimination (*a*) tells us that items may be biased and/or that items are not

assessing the same underlying trait, which would indicate there is multidimensionality in the test in question (and therefore rejected by objective measurement). Therefore, Wright (1997) also points out, item discrimination should not be treated as a parameter and that guessing (or pseudo chance) should not be treated as a parameter, but rather treated as person unreliability. Therefore, the Rasch model, is only similar to other IRT models in that it parameterizes item difficulty but that is where the similarity ends. It differs from other IRT models in that it parameterizes person agreeableness (person ability) and does not parameterize item discrimination or pseudo-chance but treats these as multidimensionality and person unreliability (Bond and Fox, 2007).

As pointed out above, the Rasch model parametises person agreeableness being assessed, commonly known as person ability. Bond and Fox (2007) point out that the 1-PL, 2-PL and 3-PL models are so called because the test in question is parameterized by one, two or three parameters but that the sample of persons are not parameterized at all and are instead characterized by a distribution. Bond and Fox (2007) state that proponents of the two and three parameter models claim that their data fit when such parameters are in place and further argue that this is hardly surprising due to fact the values of these second and third parameters are manipulated expressly for that purpose; in other words, such IRT proponents manipulate their model to fit their data. Rasch proponents have heavily criticized this 'model fits data' approach. Bond and Fox (2007), for example, state that the adding of parameters (i.e. guessing and discrimination parameters) to manipulate the model so that it fits the data 'robs the data of its fundamental measurement

properties' (p.265). The Rasch Model, instead of taking the 'model fits the data' position takes a different approach in that it attempts to make the data fit the model. Therefore, the Rasch Model is not just a way of describing the data, but rather used as a standard that utilizes particular criteria that the data must fit in order to be compliant with the model (Bond and Fox, 2007).

It can be seen that the Rasch Model has an invariance property. In other words, the model does not change and therefore the measure in question does not change. This means that the ability of the test-takers is independent of the test they take and, similarly, the difficulty of the test items are independent of the test-takers (Sick, 2008b). A psychological test that has an invariance property is a test that is stable and accurate. Therefore, when using such a test the researcher or test developer will find that person ability estimates and item difficulty estimates should not vary. In describing such invariance, Georg Rasch (1961) puts it succinctly:

> The comparison between two stimuli should be independent of which particular individuals were instrumental for the comparison; and it should also be independent of which other stimuli within the considered class were or might also have been compared.
>
> Symmetrically, a comparison between two individuals should be independent of which particular stimuli within the class considered were instrumental for the comparison; and it should also be independent of which other individuals were also compared, on the same or some other occasion (Rasch, 1961, p. 332).

What this means is that the Rasch Model is actually being used to construct the variable of interest (De Ayala, 2009). De Ayala (2009) points out that this is analogous to the physical sciences in that the Rasch model is seen to be the standard by which one creates a measure and likens it too constructing a measurement of time.

> The measurement of time involves a repetitive process that marks off equal increments (i.e. units) of the (latent) variable. In order to measure time we need to define our unit (e.g. a standard period of oscillation). With the Rasch model the unit is defined as the logit…Therefore, analogous to time measurement, our measurements are based on the (repetitive) use of a unit that remains constant across our metric. (p.19)

The Rasch approach is different to IRT and other multidimensional approaches that use different parameters to allow for the idiosyncracies in the sample of data. This is a major difference between the Rasch approach and other quantitative approaches in the human sciences that include the developing of psychological measures. A typical CTT example is that of factor analysis which is a statistical method routinely used in test development in order to identify the structure of the construct in question. The two main methods of factor analysis are exploratory and confirmatory. Both of these methods can be criticized in that they can be seen to have shortcomings in the attempt to achieve objective measurement (Marsh 1993). Exploratory factor analysis can be criticized for being designed to identify multiple constructs whereas the Rasch approach is designed to identify one construct (Sick, 2011). Furthermore, factor analysis does not allow the researcher to check in advance of any factor structure because the result is what the data produces. This is in sharp contrast to Rasch that

allows prior checking of data giving the researcher the opportunity to omit misfitting persons or items and therefore improving the data.

In the same way that exploratory factor analysis can be criticized, so too can confirmatory factor analysis (CFA). Confirmatory factor analysis does allow the testing of factor structures before any final solution and also provides various indices in order that the extent to which the structure fits the actual data can be tested. The problems with confirmatory factor analysis are that scores are attributed to participants' attributes (Stevens, 1946), the measure is directly related to the scores and parameters are sample dependent (Ewing, Salzberger and Sinkovics, 2005). Therefore, CFA does not construct the objective fundamental measurement that Rasch does. The reason for this is that the data is seen to be the model in that no single item or person data is discarded due to unreliability.

IRT approaches can also be criticized by the fact they create parameters for guessing or pseudo chance. The Rasch Model calculate *fit statistics* (fit statistics are explained more fully below) in order to show the extent to which there are misfitting persons. Misfitting persons are those persons whose scores have considerably high residuals in that their observed score is considerably different to their predicted score. Moreover, fit statistics not only address misfitting person but also address misfitting items. Similar to misfitting persons, misfitting items are those items that have high residuals in that their observed rate of endorsement is considerably different to their predicted rate of endorsement. In this way unreliable items or persons can be omitted from the data set, thus leaving a data set that has more integrity.

It can be seen, that in most quantitative methodologies, when a model does not sufficiently describe the data being analysed, it is the model rather than the data that has to change. This causes inconsistency of measures and makes it impossible to create a reliable and unique measure of the construct in question because one has to constantly change the model to fit the data. This amounts to the same as developing a weighing scale based on each case at a time which would be useless as it would not give stable and easily communicated results. Any results obtained from a different sample or different conditions cannot be compared and communicated in a reliable foundation. Wright (1999), a leading Rasch proponent, states,

> As long as primitive counts and raw scores are routinely mistaken for measures by our colleagues in social, educational, and health research, there is no hope of their professional activities ever developing into a reliable or useful science. (Wright, 1999 p.704)

Rasch is different in its approach, in that it takes on the *data must fit the model* approach, which means that a measure designed in one study can be applied to other situations. The Rasch approach therefore takes an approach that results in a measure that has integrity and stability in order to construct a fundamental measure (Bond and Fox, 2007).

The Rasch model started out as a dichotomous model and therefore used two choice items (Bond and Fox, 2007). Central to the Rasch Model is the difference between the person's ability (typically assessed by inventories) and the difficulty of the items. The probability of a specific response to a specific test item can be modelled as a

logistic function of the difference.

If a person's agreeableness or ability is high and an item is easy, the person has a high probability of endorsing an item. This is the basic idea behind Rasch analysis and which sets it apart from Classical Test Theory and Item Response Theory. However, there is a Rasch model for rating scale data (or polytomous data), commonly known as the Rasch-Andrich model, which is an extension of the dichotomous model (Andrich, 1988). This is important because in the current project a polytomous scale is used (a 5-point Likert scale). Further in this section I explain the Rasch procedure that includes its mathematical underpinnings and how they apply to the polytomous rasch model. However, first I look at how the Rasch Model related is related to validity.

The purpose of using the Rasch Model and its various statistical procedures is to enhance, as much as possible, the validity of the psychological test in question. From a Rasch perspective, this means examining the structure of test data and how it relates to the content of various test items. Providing evidence of unidimensionality (when the data fits the Rasch Model sufficiently enough) is evidence that the test has construct validity in the sense that it has internal consistency. This is in keeping with Messick's (1995) view where he proposes that measurement should refer to aspects of underlying construct and internal consistency.

The Rasch Model is uniquely adept at providing evidence for construct validity in that it can provide both empirical and statistical evidence of unidimensionality. It

does so by examining the fit of the sample data to the Rasch Model. Unique to Rasch, analyses exist that examine how items target persons on the measure and this can therefore identify if the psychological test has sufficient items to address all aspects of the measure (Bond & Fox, 2007). With regard to validity, this gives the test developer information that can give valuable information in order that the test may be improved (e.g. items added). Rasch analyses also exist that show item hierarchy via person-item maps and keymaps. Hierarchy refers to the range of item difficulty and such a hierarchy can provide valuable evidence for construct validity. For example, if the theory behind the test in question posits that spiritual experiences in therapy are rare then items assessing spiritual experiences should be more difficult to endorse. In light of this, it is arguably of great importance that initial test items are based on a well-defined theory so that such an item hierarchy can be compared to such a theory.

The Rasch Procedure

The Rasch procedure involves firstly removing persons and items for which no definite estimates of person capacity to endorse an item or item difficulty can be made, that is persons who endorse all items as well as persons who endorse no items. In rating scale data this would be removing all persons with a maximum score and all persons with a minimum score. Similarly, items would be removed which had been most fully endorsed by all persons as well as removing items which no person had endorsed. In rating scale data this would be removing items where all persons had given a maximum score and removing items where all persons had given a minimum score. The reason for removing such items and persons is that *items* must be bracketed by persons who are 'more able' and 'less able' than the item is difficult

and *persons* must be bracketed by items that are easier and harder than the person is able to endorse (Bond and Fox, 2007). In other words, for enough information to make an estimate of where a person or item is on a latent variable, some items are needed that are too easy and some items too hard as well as some persons 'too able' and some persons 'not able enough.'

The next stage of Rasch analysis involves calculating item difficulty and person agreeableness (ability). For items, this involves calculating the not-endorsed to endorsed ratio and for persons, the endorsed to not-endorsed ratio. It therefore establishes the relative difficulty of each item as well as the relative ability of each person. The greater the number of items a person has endorsed, the more 'able' that person is. The greater the number of persons endorsing a particular item, the more endorseable is that item. With item difficulty and person ability being interdependent, the Rasch model constructs a unidimensional measure simply because persons and items are on the same metric scale. Therefore, the Rasch model can be correctly thought of as a single parameter model.

Once ratios are calculated, they are then converted to their natural log odds (or logits for short). (The conversion to logits from rating scale data is a little more complex but is based on the same principles as the dichotomous model.) The rationale behind converting scores to logits is that it converts data to true interval data and it avoids the problem of compression at the ends of the raw score scale. It is also natural and convenient (Bond and Fox, 2007). The reason why logits are seen as natural is to do with the fact that they are based on e. In mathematical terms, the number e (2.718…)

is the maximum possible result when compounding 100% growth for one time period (Azad, 2009). It is the base amount of growth shared by all continually growing processes and represents the idea that all continually growing systems are scaled versions of a common rate (Azad, 2009). Therefore converting pass/fail ratios to their natural logarithm gives a figure which is not only true interval data but is also more representative of natural growth or natural development in physical or psychological sciences. Figure 4.4 shows a percentage scale along the horizontal axis with the corresponding logits on the curve. The logit curve is a more realistic representation of natural growth than a simple percentage scale. The equal increments on the percentage scale correspond to the unequal amount of logits on the logit curve. For example, one can see that the 10% difference between 50% and 60% when converted to logits is the difference between 0 and .41 (a difference of .41) but the 10% difference between 80% and 90% when converted to logits is between 1.39 and 2.2 (a difference of .81). This is arguably more representative of any physical or psychological test as one can imagine it being a much more difficult achievement to get from 80% to 90% than it would be to get from 50% to 60%. Simple percentage scores in this case would not naturally reflect the performance or growth whereas logits would.

Ability of persons and item difficulty are inextricably interrelated, and are both, in turn, related to the probability of endorsing an item. Let us say that there is zero difference (0 logits) between the ability of a person and the difficulty of an item (difficulty of an item for that particular person). The probability of that person endorsing such an item would be .5 (or 50%). In another example, let us say that the

difference is +1 logit, where the person has, for example, experienced stronger relational depth than the item measures. The probability of the person endorsing that item would be .73 (a 73% chance of the item being endorsed). Looking at it in reverse, if the difference was -1 logit, where a person has, say, experienced less strength of relational depth than the item measures, the probability of endorsing the item would be .27 (or 27%). Taking a much larger figure, if the difference between what the person experiences were much more, say +3 logits more than the item assesses, then the probability of endorsing such an item would be 0.95 (95%). If the person experiences much less than the item assesses by -3 logits the probability would be 0.05 (5%), respectively.

A distance of one logit is the increase in probability of endorsing an item, by a factor of e (i.e. the number 2.718…). A logit then, can be seen to be a certain distance along a variable just as an inch is a distance along a ruler. Therefore, we may view this logit measure as a logistic ruler and use the idea of a yardstick or physical ruler to help us understand this logistic ruler (Wright (n.d.) cited in Bond and Fox, 2007). This means that when logits are used in statistical analyses they can be used to identify different groups of persons along a construct (i.e. those with larger and smaller amounts of the construct).

Once scores are converted to logits, the mean and variance for item logit distribution are calculated (Bond and Fox, 2007). The item logit variance and person logit variance are used in a number of Rasch calculations which test for reliability and validity. Rasch analysis is not just simply one calculation but consists of a number of different analyses and statistics. The validity and the reliability of the test or scale

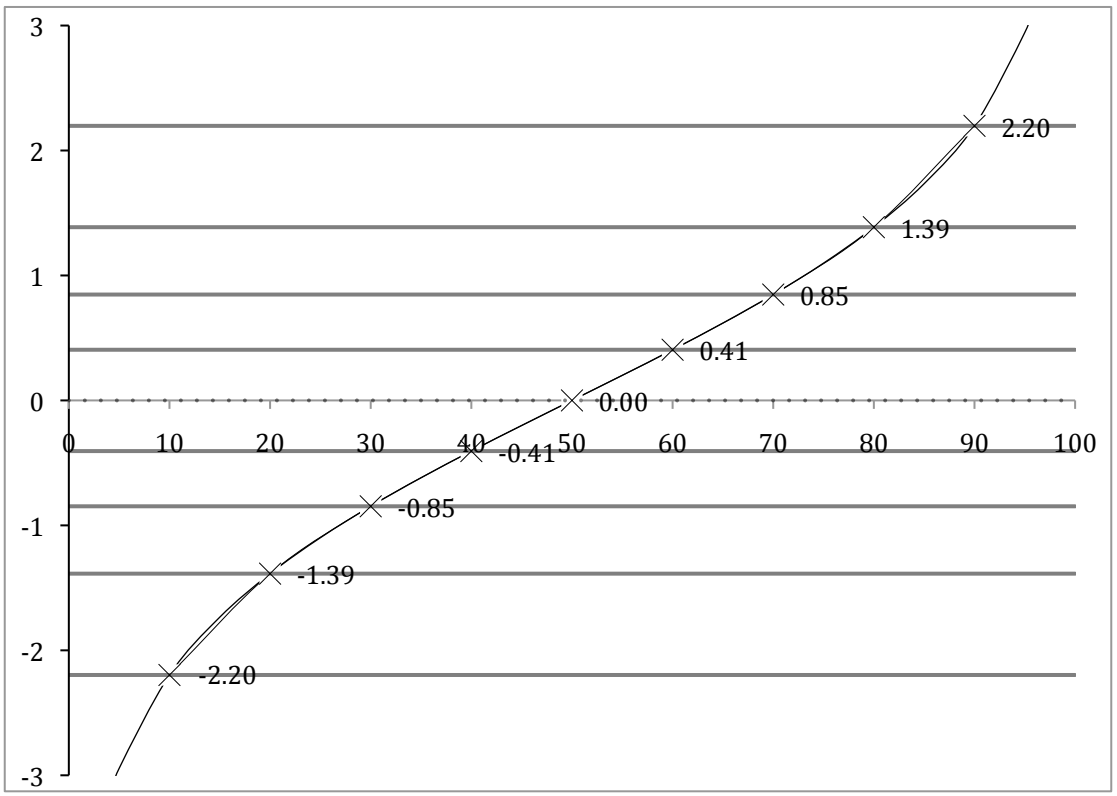

Figure 4.4. Percentage scores, from 0 to 100 in increments of 10 (across) correspond with various points on the logit curve. For example, a score of 20% when converted to a logit would be -1.39.

is tested by the Rasch model by using mathematical formulas to calculate the probability that a person will get an item correct and that an item will be answered correctly by a person (Bond and Fox, 2007). If data is different from what is predicted (from what is probable) then such data indicates that the data does not fit the Rasch mathematical model (Bond and Fox, 2007). Fit statistics are central to testing whether the data fits the Rasch model or not.

Fit Statistics

Fit statistics help the researcher identify how unidimensional each item is in that it identifies any items which do not fit the Rasch model and identifies persons which

may have assigned ‘unlikely’ scores (Tennant and Pallant, 2006). For example a typical ‘non-fitting’ person would be a person who endorses items which are particularly difficult items and seems unable to endorse those items which are particularly easy. In this later example, this may mean identifying persons who guessed an item.

With regard to misfitting items and persons, there are two misfit statistics; infit and outfit (Bond and Fox, 2007). The infit statistic is a weighted estimate that gives more weight to responses close to modeled estimates of an item (Bond and Fox, 2007). It is therefore inlier sensitive which is that it is sensitive to where most of the information is. The infit statistic is typically influenced by response patterns in the data, is hard to identify and put right, and therefore presents a threat to the internal validity of the scale (Bond and Fox, 2007).

The outfit statistic is different from the infit statistic in that it is an unweighted estimate that is more likely to be influenced by unpredicted responses, is easier to diagnose, and presents less of a threat to measurement (Bond and Fox, 2007). Outfit statistics are more sensitive to outliers (extreme scores) where less of the information is. High infit statistics, therefore, are a little more problematic than high outfit statistics. Both infit and outfit statistics have an expected value of 1 (Bond and Fox, 2007). When actually analyzing Rasch data, however, researchers typically are concerned when the mean square (MNSQ) fit statistics exceed 1.5. The higher the infit or outfit statistic, the more questionable the information (Bond and Fox, 2007).

Rating Scale Data Analysis

Another advantage of Rasch analysis is the fact that, unlike classical approaches, it is able to identify unreliable anchor points on polytomous data (e.g. rating scale data) where there are more than two response categories or anchor points (Bond and Fox, 2007). Commonly known as the Rasch Andrich Model, it was derived by Andrich (1978, cited in Bond and Fox, 2007). The Rasch-Andrich model addresses the fact that classical test methods do not recognize the subjectivity involved in choosing one response category over another (Bond and Fox, 2007). For example, often on a 5-point Likert scale, the same score of 1 to 5 is often assigned to every item's rating scale where, for example, 1=strongly disagree, 2=disagree, 3=neither disagree not agree, 4=agree and 5=strongly agree. This assumes that every item has the same difficulty in that, a score of 1 is assigned to all items for strongly disagreeing with the said item. Because of this, the Rasch model looks at the difficulty of each anchor point for each item and does by actually examining how participants responded to the item. So, for example, there may be a pattern where one particular questionnaire item is more difficult than another and a response of 2 (somewhat) on one item may correspond with 3 (very much so) on another. The RDI example below (adapted from Bond and Fox, 2007) illustrates this

Therapist was there for me	1	2	3	4	5
Therapist and I felt close to each other		1	2	3	4	5

Mathematically, the rating scale Rasch model is an extension of the dichotomous model (Bond and Fox, 2007). In practice this means that if an item has five response categories, it is seen by the Rasch model as having four thresholds. Item thresholds

are the point between two response anchor points and are usually referred to as k, and ability estimates are referred to as F (Bond and Fox, 2007). In Rasch analysis there is an ability estimate for each item threshold where the ability estimate can be seen to be the threshold where the person has a .5 chance of endorsing one anchor point over another (Bond and Fox, 2007). The distance between each anchor point is referred to as a step threshold and is the distance between each rating scale category or anchor point (Bond and Fox, 2007). These distances, unlike the classical approach, are not assumed as equal and are calculated by observing how participants actually respond to the categories of all items. If participants find it difficult to distinguish one anchor point over another then this will be apparent in the results of Rasch data analysis. In this case neighbouring anchor points will be too close in proximity suggesting that the two anchor points may need to be collapsed into one. In a similar way, if anchor points are too far apart this would suggest another anchor point is needed. As well as this, Rasch analysis will establish the common pattern to all items on the scale thus informing the researcher of what, if any, anchor points need to be collapsed.

Reliability Indices for Persons and Items

It is important to note that all the various indices and coefficients described below have one thing in common and that is something called separation. Separation in Rasch analysis refers to the separation or the spread of the sample (Bond and Fox, 2007). The sample can be the sample of items or the sample of persons. Either way, separation statistics of persons or items involves looking at how the spread of items or persons are distributed and how they actually perform (Bond and Fox, 2007).

This means that Rasch analysis, unlike its classical equivalent, has the ability to examine a measure in detail by identifying different groups of persons and different groups of items across the measure as a whole and whether the measure in question sufficiently targets appropriate samples of persons.

The Rasch model provides a number of different forms of index score which give us an indication as to how well the items on a measure vary enough to be able to assess persons of differing ability (Bond and Fox, 2007). It therefore examines how well spread the range of items are in relation to the population they are measuring. It does something similar for the persons in a sample in that it looks at how reliable the sample of persons are. In other words it gives us an indication as to whether the same persons would be just as able or unable if they were given another measure which assessed the same construct. Rasch reliability is equivalent to the classical 'test' reliability (such as Cronbach's Alpha) but provides reliability indices for both persons and items. Also like its classical counterpart, values are between -1 and 1, with values closer to 0 indicating low reliability and values closer to 1 higher reliability. However, classical reliability testing is really a "person sample reliability" rather than a "test reliability" because it is based on the "true person variance / observed person variance" (Bond and Fox, 2007). Unlike its classical reliability, which typically gives an internal reliability of the test, Rasch reliability gives an estimate of replicability of person placement on the measure and item placement within a hierarchy of items (Linacre, 1997). The index value pertaining to *item reliability* is designed to indicate the replicability of the logistical positions of persons if they were given a measure assessing the same construct (Bond and Fox,

2007). A low value of item reliability index would indicate a small range of persons (e.g. a narrow range of ability or a similar amount of the construct being assessed). Linacre (2011) suggests that to improve item reliability persons with extremely high and extremely low amounts of the construct should be targeted.

Item reliability index

Item reliability is a similar index to its person counterpart in that it gives an indication as to the extent to which we can expect the same set of items to perform given another sample of persons which had the same range and ability to endorse such items (Bond and Fox, 2007). The Item Reliability index tells us how stable the items are or how stable their estimates are. Would item x still be more difficult as item y?

Separation Indices. A separation index in Rasch analysis can be seen to be another form of reliability index and is sometimes referred to as a reliability separation index or coefficient (Linacre, 2011). However, not to be confused with the reliability indices mentioned above, separation indices are, for items, an estimate of the spread or separation of items on the measured variable or, for persons, the spread or separation of persons on the measured variable (Bond & Fox, 2007). It can be defined as the number of statistically different performance levels that the test can identify in the sample. Therefore, a person separation index is based on the number of statistically different levels of persons that are identified by the range of items and an item separation index is based on the number of statistically different levels of items that can be identified within the range of items (Bond and Fox, 2007). Mathematically, the separation index is commonly known among Rasch researchers

as G and is the adjusted standard deviation (person or item) divided by the average measurement error. Where G = Separation index = "True" standard deviation / Average measurement error (Linacre, 2011).

Item Hierarchy

Unique to Rasch is the fact that it can calculate the hierarchical order of item difficulty which enables the construction of a measure that is comparable to the construction of a ruler with equal increments (Linacre, 2011). In other words items are calibrated and are able to be plotted on various strengths of the latent construct. Therefore, Rasch analyses can aid the test developer to identify which items assess which strengths of the construct. In turn, this aids the test developer to investigate the integrity of the test and facilitates and guides improvement to validity enhancement.

Item and Person Strata.

Linked to item hierarchy and similar to G mentioned above, there is another method which identifies spread of persons or items and is called strata, H, and is computed as: H = (4G + 1)/3 (where G = Separation index = "True" standard deviation / Average measurement error) (Linacre, 2011). Wright and Masters (1982) state "... if we define statistically distinct levels of item difficulty as difficulty strata with centers three calibration errors apart, then this separation index G can be translated into the number of item strata defined by the test H=..." (p. 92). Each strata or level is located three standard errors apart. A distance of one standard error is more than 1.96, which happens to be the distance corresponding to the .05 significance level. Just like G, a strata index can be used to identify statistically different levels of item difficulty and

person agreeableness that are identified by items or persons on the measure. However, G is often seen to be a much more conservative index than H. Wright and Masters (1982) warn that it is difficult to know whether observed differences in the sample are real or down to measurement error. They add that if the true standard deviation of a sample was the same as the measurement error then G would be equal to 1. This would make the test reliability to be 0.5, which may indicate down to chance alone.

When identifying strata with either persons or items, Elliott et al (2006), in a study that examined the validity and reliability of the Symptom Checklist-Revised, illustrated how that there does not necessarily have to be a precise analysis when identifying strata. In their study, they conducted an informal qualitative analysis in order to obtain a better understanding of their strata (aided by using separation statistics and keymap measure step considerations). They did so by categorising items with similar difficulty, which resulted in 11 partially overlapping qualitative groups of items. In the current study, I intend to follow this example for both items and persons but with some exceptions. The exceptions involve accuracy; it is possible to identify distinct strata by producing various outputs from the Rasch software (i.e. Winsteps). For identifying item strata, a Rasch table can be produced which shows all items in measure order. In other words, for each item, a measure value is given in logits. This value is the estimate or calibration that indicates item difficulty. Because the standard error is also given, one can easily work out which items, or groups of items, are a standard error apart from other items. Producing a keymap (from the Rasch software) also helps considerably to identify strata because

items are physically mapped in difficulty order. What a keymap does, therefore, is actually map the measured variable by plotting persons against the response categories of each item. This can enable the categorizing of groups of items according to participants endorsement of, in the case of this study, relational depth questionnaire items. This should identify and distinguish different groups of items according to their difficulty.

For persons, a person-item map is a very useful tool in identifying strata of persons. Person-item maps can be thought of as a ruler that has measurements of participants' capacity to endorse an item on one side, along with the items that need to be endorsed in order to measure such person capacity on the other side (Linacre, 2011). Person 'capacity', or in the case of the current study, participants' strength of relational depth, is referred to as the person measure, and item difficulty required to assess such person capacity is called the item measure. This mapping of persons and items together enables the researcher to see how well their items target participants in their research. For example, researchers are able to see whether the 'most able' participants are targeted appropriately by items. Therefore, not only does a person-item map enable the investigation of person strata but also visually demonstrates how well the range and position of the person distribution compares with the range and position of the item distribution.

Figure 4.5 is a simplified example of a person-item map and is not based on any real data. In this example, mean item difficulty can be seen to be zero. Persons 2 and 3 have estimated measures pretty much on the zero point. The same can be said for

items 2 and 8. What this means is that these persons and items are at the same point on the map and because of this we can make a prediction that these persons will endorse at least one of the two items. Similarly we can predict that at least one of the items will be endorsed by at least one of the two persons. Looking at other examples, persons 2 and 3 are more than 50 percent likely to be successful at items 3, 1 and 7 because these items are easier which is indicated by the fact they are lower down the item map. The same two persons are less than 50 percent likely to be successful at items 4 and 6 because these items are more difficult which is indicated by the fact they are higher up the item map. Coming back to person strata, one can see that items on the right tend to form in clusters with each cluster separated by the grey line (for illustration purposes). In reality, clusters are not always as easy as this to identify but this illustration gives some idea.

With the exception of the person-item map, other examples of the maps and tables mentioned above are not reproduced here but will be produced in the results section and full explanations given regarding their interpretation as to how they display results and data from the research here (please see Chapter 7).

Rasch Principal Components Analysis (PCA)

Another tool that Rasch has in its toolbox for assessing the validity of a measure is principal components analysis (PCA) of the Rasch residuals (Linacre, 2011). PCA is usually associated with classical test theory. However, there is a Rasch equivalent to classical PCA but unlike its classical counterpart, the Rasch version analyzes only the variability in the set of items after the linear measure has been accounted for (Bond & Fox, 2007).

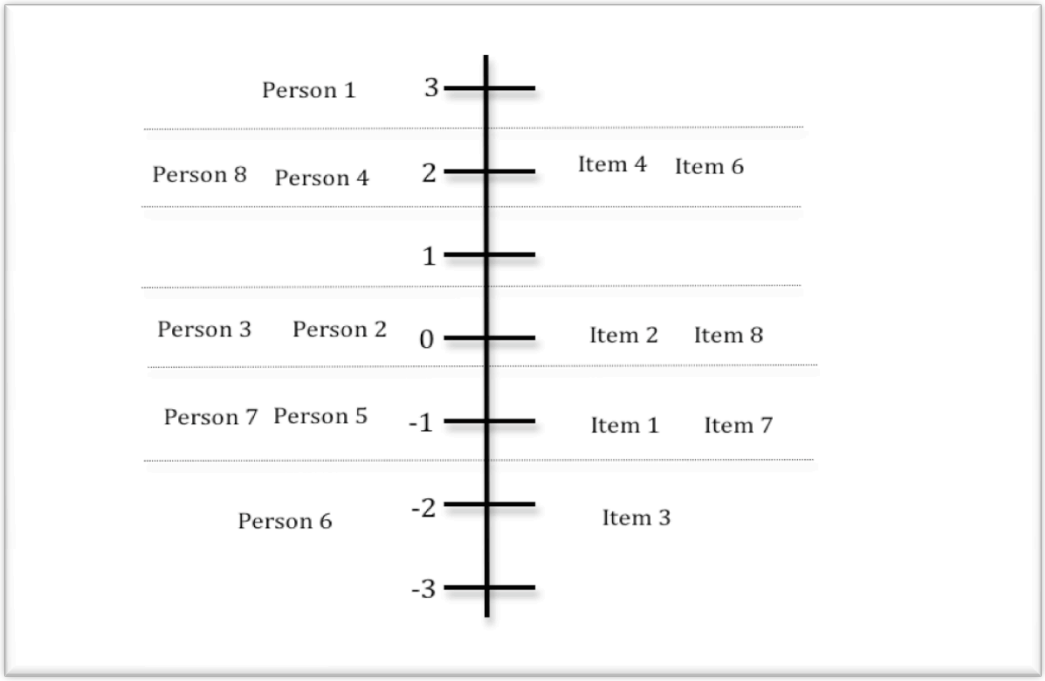

Figure 4.5 An example of a person-item map. Adapted from Sick (2008a)

Therefore, in Rasch PCA, it is the residuals that are analysed and explored after the primary dimension has been removed. That is, the measure has been statistically removed and what is being factored is what's left over, which is the misfitting component of the variance. Residuals in Rasch analysis are values that are essentially the difference between the Rasch model's expectations and the actual performance (Bond & Fox, 2007). If a measure is truly one-dimensional we would expect any residual factor loadings to be small and not provide any meaningful factor. Therefore by not finding any meaningful secondary factor we can be reasonably sure that our measure possesses 'psychometric unidimensionality' (Linacre, 2011). Unlike any classical counterpart, Rasch PCA has an advantage in that it enables any secondary factors to be assessed for size and meaning. For example, it may be that secondary factors are easily detected but not really large

enough to cause a threat to the unidimensionality of the measure. Such secondary dimensions will need to be considered carefully and, for example, may be seen to be an unrelated factor of a similar dimension to the measure in question. If this is the case Linacre (1998) suggests that diagnostic action be considered such as grouping the items into subtests and constructing additional latent variables .

4.2.9 Summary to Reliability and Validity

It is evident that establishing evidence that a psychological test has reliability and validity is a complex issue. Further, there is no consensus of opinion as to how one should obtain this evidence or which methods to use. In light of the above reviews of Rasch analysis and how it compares with other quantitative approaches, I approached this research project as a proponent of the Rasch Model. This is due to the fact that I wish to create and develop a measure of relational depth that is based on the fundamental principles of Rasch philosophy mentioned above. This means that I took on the position of the 'data must fit the model' approach. However, some CTT approaches have been used initially to create an initial relational depth inventory but later analyses use Rasch extensively.

4.3. Data Collection methods

4.3.1. Introduction

This section focuses on the varying methods of data collection but specifically focuses on the methods of data collection used in this research project namely web-based data collection and practice-based data collection. Regarding web-based data collection, various studies are reviewed which specifically look at various aspects such as perceived privacy, anonymity, self-disclosure and social desirability.

Typically such studies compare web-base data collection with pencil and paper data collection in order to find out whether respondents behave differently between these two methods of data collection.

Regarding Practice-based research, literature is reviewed which compares the evidence-based practice paradigm and the practice-based paradigm. This is in order that practice-based research is put into context.

4.3.2 Web-based Data Collection and Research

4.3.2.1 Introduction

With the use of email and the World Wide Web, the internet is increasingly becoming a popular research tool for research and data collection. Most of us who are familiar with email and used to surfing or browsing the web have probably been at some point asked to complete an online survey. Many of these are probably from market-research companies. However, there has seen somewhat of a revolution as social scientists and many other researchers seem to be paying more attention to the use of the Internet as a data collection tool. It would seem that in many cases traditional paper-and-pencil type research questionnaires are being replaced by online recruitment, especially for large-scale data collection in a relatively short space of time. As well as providing fast access, online questionnaires and surveys provide a reliable way to collect large amounts of data easily. Despite the fact that online research began during the mid 1990s (during the 'internet boom') it is only more recently that creating and conducting online surveys has become easier. Familiarity with web programs, HTML code and scripting programs is no longer needed as many survey software packages make designing a web-based

questionnaire easier and faster in that it no longer requires the researcher to have knowledge of these (Wright, 2005).

Hewson (2003) states that Web-based participants have the potential to enjoy a more private research experience, and therefore be more comfortable disclosing information. This would imply that internet-based research is seen as less intrusive, as it enables the participant to complete forms in their own time, potentially in their own home, without having to physically face the researcher. This implies that participants may be more likely to disclose information about themselves.

Another obvious advantage of Web-based research is the reduction in cost when compared to pencil-and-paper surveys as it eliminates the need to print large amounts of paper questionnaires and also saves on shipping costs. In addition, many Web-based surveys contain in-built software that allows raw data to be directly transferred to statistical packages, such as Microsoft's Excel and SPSS, thus saving the researcher the laborious task of manually typing in data from paper questionnaires. Furthermore, advantages of Web-based surveys also include being able to access potential respondents form virtually anyone in the world, as long as they have a device that can access the Internet.

However, there may be disadvantages and limitations to Web-based data collection, and researchers who are intending to use the internet need to be aware of some of the potential difficulties. Hewson (2003) states that internet samples may be biased in that research has found that it is mainly technologically proficient, educated, white,

middle-class, professional males who would utilise web-based research, although he adds that this is changing as the internet grows. Other problems with internet samples include issues such as controlling for multiple submissions, incomplete responses and unacceptable responses (Schmidt, 1997). Wright (2005) states that although problems are inherent whatever data collection method is used, there are some that are unique to Wed-based design such as sampling issues in that little is known about characteristics of online communities. Other problems could include the issue of multiple responses where it could be difficult or impossible for the researcher to know if a respondent completed a Web-based questionnaire twice, especially if the online data collection is truly anonymous where neither name, email address or Internet Protocol (IP) addresses are recorded by the administrator.

In light of the fact that the current project involves the use of both Web-based data collection and paper and pen data collection, it is appropriate to look at the current research that focuses on comparing these two methods. In particular, I wish to document research that focuses on respondent perceptions of confidentiality and privacy, possible differences in disclosure and differences in data reliability. This is mainly because the current study utilizes Web-based questionnaires and paper and pencil questionnaires to ask about clients' experiences of therapy and therefore collects private, potentially sensitive information. In addition, one of the current research studies utilizes both a Web-based survey and a PP survey where data from each of these administration methods are used as one sample in a data analysis. It is important to establish whether responses to web-based questionnaires are comparable

to their pencil-and-paper equivalents. Therefore, literature that looks at the fidelity of online data collection versus traditional paper and pencil will also be highly relevant.

There are many research studies that focus on computer-based data collection versus pencil-and-paper data collection. In such studies, most of the computer-based studies are not also Web-based in that computers used in the research were not connected to the internet and questionnaires were simply presented electronically using the computers software. There is therefore a vast amount of literature which looks at computer-based studies. In light of this, for space reasons you are being selective and focusing on only the most relevant studies.

4.3.2.2 Perceived Privacy, Anonymity and Self-Disclosure.

Perhaps one of the most crucial issues when completing a questionnaire that is presented on a website is the concern as to whether the information that is disclosed is anonymous and confidential. Moreover, what, how and how much we disclose may depend on our perceptions of anonymity and confidentiality.

Van de Looij-Jansen and de Wilde (2008) studied adolescent pupils (14 to 16 year olds), administering a questionnaire that enquired about private information, including conduct problems, peer problems, health risk behaviour (such as alcohol consumption, smoking, cannabis use and whether they had ever engaged in sexual activity). Both the Web survey and the PP survey were presented in class and each pupil completed one or the other. No significant differences were found in perceived levels of privacy and anonymity between those who completed PP questionnaires and those who completed Web-based questionnaires.

In a similar study, Denniston et al (2010) administered the Youth Risk Behaviour Survey (YRBS) to ninth and tenth graders (15 to 16 year olds) in the USA. The YRBS essentially enquires about risky behaviours such as tobacco use, alcohol consumption, unprotected sex and unhealthy eating, and again enquiring about potentially sensitive and private information. Both Web-based surveys and PP surveys were presented in-class. The respondents who completed their study in the Web-based condition (compared to the paper-and-pencil condition) were significantly more likely to feel unsure of their anonymity, more likely to feel they wished they had completed the survey in a more private place and more likely to think that taking the survey on the internet was not private. However those in the Web-based condition were less likely to be concerned about being identified when they visited other sites on the internet. This could imply that there may have been some aspect of the particular study that caused respondents to feel unsure about anonymity.

In other research, Hanna, Weinberg, and Dant et al (2005) investigated how anonymity (anonymous versus identified) may interact with method of administration (Web versus paper-and-pencil). In the study, business undergraduates were asked about their private lives. There was no significant difference between web and paper responses in the number of sensitive questions answered or the amount of self-disclosure. Furthermore, there were no significant differences found between anonymous and identified responses. Interestingly, an interaction was found between method and anonymity; for the paper condition self-disclosure was slightly higher for the identified respondents than the anonymous respondents, but in the web

condition the opposite was true: self-disclosure was higher in the anonymous respondents than the identified respondents. This indicates that people may be more comfortable disclosing information if they feel it is anonymous.

In another study which also looked at self-disclosure, Joinson, Woodley and Reips (2007) found that non-disclosure to a sensitive question (about earnings) was significantly higher when participants were asked to use a secure log-on web page procedure (using personalized passwords) rather than on a personalized URL (which included an encoded identifier not obvious to the respondent). This again supports the idea that it is perhaps perceived anonymity that determines levels of self-disclosure.

4.3.2.3 Social Desirability.

Another factor concerning internet surveys concerns demand characteristics and social desirability. Socially desirable responding has typically been conceptualized as consisting of two components: impression management and self-deceptive enhancement (Paulhus, 1991). Impression management is the deliberate tendency to over report desirable behaviours and under report undesirable ones. Self-deceptive enhancement is the tendency to give honestly believed but overly positive reports about oneself. The question is whether responding via an internet survey as opposed to paper-and-pencil, reduces social desirability.

Joinson (1999) conducted a study where undergraduates were recruited during class to complete questionnaires that enquired about social anxiety, social desirability and

self-esteem. There was a Web-based condition and a paper-and-pencil condition. Both conditions were completed in class but none of the participants in each condition knew about the participants in other condition. Within each condition there was an anonymous sample and an identified sample. Findings were that undergraduates using Web-based questionnaires, as opposed to paper-and-pencil questionnaires, reported lower social desirability, lower social anxiety and marginally higher on self-esteem.

Risko, Quilty and Oakman (2006) conducted research that specifically focused on social desirability when responding on the Web. Here the researchers administered a battery of three social desirability questionnaires to a student sample and to a self-selected sample. The student sample was drawn from undergraduate psychology students and was what the authors describe as a 'typical captive student sample' where active recruitment took place; students received credits toward their course. Here, one half of the student sample completed paper-and-pencil and the other half completed Web-based questionnaires (as part of laboratory experiments with psychology undergraduates). The self-selected sample was obtained via the posting of links to two websites that housed online psychology experiments. Any person interested would be able to select any study they wanted. The self-selected sample was only Web-based (there was no paper-and-pencil participation) and was therefore collected passively in that there was no active recruitment of participants. No significant differences were found between Web-based respondents and paper-and-pencil respondents for reported social desirability, in any of the questionnaires. However, there was a difference between student web conditions and self-selected

(web) conditions in that the self-selected (web) condition scored lower on self-deceptive enhancement. The authors suggest that this difference indicates greater egoistic bias or need for social status and prominence in the student sample.

In a Dutch study, Van de Looij-Jansen and De Wilde (2008) found that Dutch adolescent pupils in a paper-and-pencil condition reported generally more prosocial behaviour, than their counterparts in the Web-based condition. In addition, in a specific question about carrying weapons (considered to be an anti-social behaviour), pupils in the Web-based condition reported significantly more often carrying a weapon than the paper-and pencil condition. The authors do not discuss this particular finding about carrying a weapon but I would suggest this implies that participants felt more anonymous in the Web-based condition than the paper-and-pencil condition.

4.3.2.4 Participation and Missing Data.

Participation is perhaps one of the most difficult aspects to monitor with regard to administration of Web-surveys. If links to surveys are sent in emails, it is virtually impossible to know that the potential participants have received such links, especially as spam filters may automatically delete emails. In addition, in placing advertisements for recruitment onto other websites, it is more often not clear how many people will have seen the advertisement and so not possible to know proportions of participation versus non-participation. However, it is possible to calculate missing data due to participants not responding to questions.

Denniston et al (2010) found that missing data (failure to answer a question) was higher in the web condition that the paper-and-pencil. However, in another study Denscombe (2006) found no difference in completion rates or missing data between Web-based questionnaires and paper-based questionnaires. In another, Raat, Mangunkusomo, Landgraf, Kloek and Brug (2007) found that more missing answers in their paper version of a youth health questionnaire than its web version.

4.3.2.5 Data Reliability

Perhaps the most important issue of Web-based collected data concerns whether it differs from PP collected data. This is important because it is advantageous to be able to generalize any research findings across the two different samples of research participants. Joinson (1999) argues that behaviour on the Internet differs from similar behaviour in the real world. If this claim is true one would expect differences in the characteristics between Web-based collected data and PP collected data. This is a most important issue if one is to generalize between web and paper conditions within one study. Research that addresses the reliability of Web-based data and paper-and-pencil data collection, generally looks at characteristic differences between Web respondents and PP respondents.

Herrero and Meneses (2006) administered two psychological distress measures, the Perceived Stress Scale (PSS) and the Center for Epidemiological Studies-Depression Scale (CESD) to 530 undergraduates. Each scale was presented in two formats; online and paper format. Participants were randomly assigned to formats. Results found that mean item responses were virtually indistinguishable between the samples

with only one item, which assessed confidence to handle personal problems higher for internet respondents than those using traditional paper questionnaires. Distributional characteristics (skewness and kurtosis) were found to be similar across groups. Internal consistency was very similar across formats as were factor structures, indicating that psychometric qualities were similar across formats.

In a similar study, but with children, Raat, Mangunkusumo, Landgraf, Kloek and Brug (2007) administered the Child Health Questionnaire (CHQ) to over a thousand teenagers in a randomized design to compare paper versus online format. Both modes of administration demonstrated comparable scale scores and evidenced comparable psychometric properties, with effect sizes of differences negligible to small (< .01 to .21) between paper and online respondents, for all subscales. Internal consistency was also good to excellent for all subscales (from .69 to .92).

Vajello, Mananes, Comeche and Diaz (2008) conducted an experiment where students completed the Symptom Check List (SCL) – 90 and the General Health Questionnaire (GHQ) -28. Each student completed both questionnaires twice; first in one format (internet or paper) and then in another (paper or internet). Their results found no statistically significant differences between formats in the majority of subscales in both questionnaires. Correlation between formats was high as was internal consistency for each questionnaire. The only difference between the two formats was in the social dysfunction subscale of the GHQ-28 where higher scores resulted when applied via the internet compared to paper. This was regardless of whether the web version was completed before or after the paper version. The

authors state that this may be something to do with the participants' interaction with the computer and that further research would be needed to explain this.

Much more recently, Campos, Zucoloto, Bonafe, Jordani and Maroco (2011) administered three inventories which assessed burnout to 150 Brazilian undergraduates. A within-subject cross randomized and counterbalanced design was used. This meant that during one week half the participants were randomly assigned to the paper-and-pencil versions and half the online versions. During the second week of the study the two groups were switched (those who had previously completed the online version now did the paper-and-pencil version and vice versa). For the online version, participants could complete them wherever they wished and whenever they wished within a specific time limit; they were not required to complete them in during lectures. The paper-and-pencil version was administered after class hours by a researcher who then left the room to allow completion. Results found no significant effects in burnout scores due to the order of administration (whether paper-and-pencil version was administered before the online version or vice versa). Burnout scores were higher in the second administration but were explained by authors as due to learning effects. Campos et al. (2011) also reported that reliability for the three burnout inventories was good and that similar high reliabilities resulted for the large majority of subscales. However, in the paper-and-pencil format some subscales showed higher reproducibility and concordance than the online equivalent. Campos et al. (2011) put this down to researcher-subject interaction during the administration of the paper questionnaires as well as a researcher presence effect versus total anonymity in the online format. Other

findings were that construct related validity did not seem to suffer due to mode of administration; factor analyses showed factorial invariance between the two formats.

4.3.3 Practice-based research

One cannot fully describe or explain practice-based research without first looking at two paradigms: the evidence-based practice and practice-based vidence paradigms. There is more to these paradigms than the names alone would suggest. They are paradigms and therefore each has a particular theory underpinning it.

Essentially, the evidence-based practice paradigm has at it roots efficacy research and this in turn is based within a natural sciences paradigm (Barkham & Mellor-Clark, 2003). This means that efficacy research focuses on the effects of psychotherapy under strictly controlled conditions. This also means that evidence-based practice is therapeutic practice that is influenced by research evidence that could be seen to be more positivist in its approach. Typically such evidence would come from research that utilizes Randomised Control Trials (RCTs) and systematic reviews. RCTs are studies 'in which people are allocated at random to either an intervention or control/comparison group. The effects of the intervention are determined by comparing the outcomes of both groups' British Association for Counselling and Psychotherapy (BACP, 2011). Systematic reviews, on the other hand,

> aggregate the findings of similar types of study addressing the same type of question, thus providing robust findings based on large amounts of data.

> Systematic reviews of RCTs, often known as meta-analyses are viewed as the most reliable type of evidence on which to base clinical and policy decisions. (BACP, 2011).

In sharp contrast, central to the practice-based paradigm is research that has been described as 'studies which use pre- and post- measures (such as CORE) to study the effects of an intervention in a particular cohort of clients, without the use of a control group. Some types of case study and qualitative research can also fit within this category' (BACP, 2011). Practice-based research studies the effectiveness of therapy in routine practice but in ways that are less elaborate (Barkham and Mellor- Clark, 2003). Therefore, the evidence-based practice and practice-based evidence paradigms are essentially distinguished by the respective underlying research approach. Put more simply, the evidence-based approach is underpinned by a more narrowly scientific approach and the practice-based approach by a less scientific approach (but arguably more ecologically valid).

This puts practice-based research into context and in light of the above, it would seem that it is perhaps slightly inferior to the other methods of research mentioned and that RCTs are more methodologically sound due to their control condition. However, Margison et al (2000) state that RCTs are not without problems in that there are 'technical limitations such as differential attrition rates, non-comparability of comparison groups, psychometric problems with outcome measures, inconsistency of treatment delivered and contamination by other treatments in trials of long-term therapy' (p. 123). Margison et al (2000) further state that most importantly to

clinicians is the poor success of RCTs predicting outcome and the consistent finding (by RCTs) that most treatments are better than no treatments. These authors also add that this is in comparison given that fact that 'the best predictor of outcome is initial therapeutic alliance' (Margison et al 2000, p.123).

There are many arguments in favour of practice-based evidence. Thomas, Stephenson and Loewenthal (2006) state that it gives a voice to practitioners and service users in that they have first hand knowledge of what works and what needs to be changed. Margison et al (2000) also praises practice-based evidence describing it as the gathering of 'good quality data from routine practice' (Margison et al, p.124).

Although practice-based research may be criticized for not being scientific, it appears that it is a valid approach to collect research data. Indeed, it is argued here that it is a more naturalistic way to conduct research having an ecological validity that its evidence-based counterpart lacks.

4.3.4 Summary of Data Collection Methods

It would appear that the data collection method may well have an effect on how a participant views their privacy and thus how they respond to personal questions when answering questions via a website compared to paper-and-pencil. However, the evidence is not entirely clear given that the various studies give different results. Regarding social desirability, it would seem that those who respond to questions presented via web compared to paper-and-pencil either report lower social desirability or there is no difference. However, there is evidence that web-based respondents report lower social anxiety and higher self-esteem and are not less afraid

of admitting to antisocial behaviour compared to paper-and-pencil respondents. Regarding missing data, the evidence is mixed suggesting there may be no difference between web and paper-and-pencil administration. There appears to be no evidence that suggests data reliability is any different for web-based collected data compared to paper-and-pencil collected data. Finally, the evidence would suggest that practice – based collected data has many advantages over RCTs. It is more cost-effective and, importantly, more ecologically valid.

4.4 Significant Events in Therapy

4.4.1 Introduction

This section of the literature review looks at significant events in therapy research. Such research can be seen as an area separate to most other research in counselling and psychotherapy. Timulak (2010) describes significant events research as a particular approach to studying client-identified important moments in that it is a type of process research that uses the client's and/or therapist's description of the event as well as enquiring about their reflections of the event. There is an idea behind carrying out this kind of research; that such important moments in therapy may well be some of the most fruitful moments in the whole process of therapy (Timulak, 2010).

Probably the first to research events in therapy and the first to coin the term 'therapeutic events' was Berzon and colleagues (Berzon, Pious & Farson, 1963) where both therapists and clients were asked to write reports about those events in group therapy they thought were important.

Sidney Bloch and colleagues (Bloch, Reibstein, Crouch, Holroyd & Themen, 1979) also studied events in group therapy and criticized Berzon et al's (1963) method for having 'insufficient definition and specification of criteria for assignment of the factors' (p.257). In an attempt to improve the methodology for classifying events, Bloch et al (1979) used a 'most important events' questionnaire then used a more structured methodology to assign events to particular classification factors.

Later, Bloch and Reibstein (1980), studied clients' and therapists' perceptions of therapeutic factors in group therapy. Reports of events were assigned to 10 sets of therapeutic factors then each of such factors further assigned to one of three basic classes: behavioural, cognitive and affective. Behaviour factors emphasized action and were self-disclosure, learning from interpersonal actions and altruism. Cognitive factors stressed understanding and were self-understanding, vicarious learning, guidance and universality. Affective factors stressed emotions and were acceptance, installation of hope and catharsis. Results found that there was a trend for therapists (compared to clients) to see behavioural factors as more important and cognitive ones as less important.

However, the pioneer of significant events research in individual therapy is Robert Elliott (e.g. Elliott, 1985) and also the first to coin the term 'significant events.' Elliott uses this term to apply to important events, helpful events, important moments and helpful moments and therefore generating an umbrella term that helps broaden the range of events research. Since the mid-eighties other prominent researchers

such as Timulak (e.g. 2010) have added to the now considerable body of significant events research.

Methods used in significant events research typically involves asking a client what was significant (important or helpful) about the particular session. The Helpful Aspects of Therapy Form (HAT; Llewelyn, 1988) is often administered immediately after a therapy session and records the significant event (either helpful or hindering). Events are then reflected upon by the client who is asked to make written comments.

In light of the above description of significant events in therapy, the current project can be thought of as coming under the general umbrella of significant events research. This is due to the fact that in all studies in the current project, participants were asked to describe an important event during the last few sessions of therapy and then further asked to reflect on that event by rating it using a version of the Relational Depth Inventory (RDI). This section therefore firstly looks at significant events research generally, such as how significant events have been classified, and then moves on to issues particularly relevant to the current project, such as significant events outcome research and significant events research that looks at relational factors.

4.4.2 Classifying significant events in therapy.

One of the first studies to identify types of significant event in individual therapy was carried out by Elliott (1985) who identified helpful and nonhelpful (not helpful but not hindering either) events in therapy. Specifically, the study's basis was on the immediate therapeutic impact experienced by the client due to counsellor responses.

Using a quantitative Helpfulness Rating Scale, clients rated their therapist's response as to how helpful it was. From a total of 86 helpful events, the study found two superclusters of helpful events; a 'Task' supercluster and an 'Interpersonal' supercluster. The Task supercluster comprised four clusters; 'New Perspective,' 'Problem Solution', Clarification of Problem' and 'Focusing Attention.' The Interpersonal supercluster consisted of four clusters; 'Understanding', 'Client Involvement', 'Reassurance' and 'Personal Contact.' Most relevant to the current project is the Personal Contact cluster with the Interpersonal supercluster: Elliott (1985) states that 'Personal Contact events involved moments in which the student experienced a person-to-person relationship with the counsellor' (p.311). Interestingly none of the other Interpersonal clusters refer to moments indicating an experience of connection (compared to other interpersonal experiences) that was momentary in nature.

With regard to the client-therapist relationship, Moreno, Fuhriman and Hileman (1995) found that clients reported that experience, insight and relationship were often the reason significant events had an impact.

Recently, Castonguay et al (2010) similarly looked at helpful and hindering events in therapy. Funded by the Pennsylvania Psychological Association (PPA) and the Committee for the Advancement of Professional Practice of the American Psychological Association (CAPPAPA), the study was the first of its kind in that it was the first process study conducted in the context of a Practice Research Network (PRN). (PRNs are essentially studies that involve collaboration between researchers

and practitioners.) Between them, thirteen therapists from different theoretical orientations saw 121 clients over a period of 18 months. Both clients and therapists described helpful and hindering events by completing the Helpful Aspects of Therapy (HAT) (Llewelyn, 1988). The most frequent events for clients were self-awareness, problem clarification, problem solution, self-insight and alliance strengthening. For therapists, the most frequent events were self-awareness, alliance strengthening, problem clarification, and positive self. As well as investigating helpful or hindering events the study also looked at the focus of significant events and found that for both client and therapist helpful events, issues related to therapy had the highest mean rating. For example, the study reported that a client had reported a helpful event as 'the sharing of responsibility of change' (p.341) thus again emphasizing the importance of the relationship between client and therapist.

In identifying client-identified helpful events, Timulak (2007) took a different approach and used qualitative meta-analysis to identify core categories. Timulak (2007) states that the study's aim was to 'go beyond structured summaries and evaluate the possibility of using a rigorous qualitative meta-analytic method to assess findings from qualitative studies' (p.305). Inclusion criteria for studies were that the studies had to be qualitative, discovery orientated (without the use of a *priori* categories), the events had to be helpful and significant, and the event had to have a helpful impact on the client. The focus of the sorting of studies was on the impact of the client. Nine core categories emerged a) awareness/insight/self-understanding; b) behavioural change/problem solution; c) empowerment; d) relief; e) exploring feelings/emotional experiencing; f) feeling understood; g) client involvement; h)

reassurance/support/safety; and i) personal contact. The most relevant core category to the current study is that of personal contact. Timulak (2007) states that out of the studies analysed, personal contact was identified as both significant and helpful in three out of seven. Typical client experiences were of the client experiencing the therapist being not like a doctor but rather like a (close) fellow human being and feeling more connected. In one study (Elliott, 1985), one client said 'It was like we weren't in the session anymore; like we were done and were chit-chattin' and he was telling me about the way came from his problem to where he is now. He was trying to make himself human. He wasn't a doctor anymore.'

Some significant events research focus on one type of event. Timulak and Elliott (2003), for example, focused on significant events that were considered empowering. Data was from a larger collection of client-identified significant events with parts of client and therapist Brief Structured Recall interviews. Twelve significantly empowering therapy events were identified from nine depressed clients. Five types of empowerment events emerged from the data: poignant, emergent, decisional, determination and accomplishment. In each of the event types, descriptions were provided for client processes, therapist processes and interaction, and therapeutic impact. Evidence suggested that the therapist's communication of empathic understanding and empathic affirmation of client growth promoting aspects contributed to client empowerment. This further evidences that client-therapist variables and especially the therapeutic relationship may contribute to empowerment and positive outcome.

In a review of significant events research, Timulak (2010) provides an overview of studies conducted, methodology used and findings. The findings of this review were wide and varied but relevant to the current project, Timulak (2010) found that more studies focus on helpful events than non-helpful events. For helpful events, those surrounding *reassurance, feeling understood* and *personal contact* contributed to the therapeutic relationship and those concerning *insight, relief, behavioural change, new feelings* and *empowerment* contributed to in-session outcomes. Differences as to what is important in therapy differed between clients and therapists; therapists seemed to prefer events surrounding therapeutic work such as insight whilst clients seem to place more emphasis on the relational aspect of therapy. Also, concerning the client-therapist relationship, findings indicated that when working alliance was higher there was an increase likelihood of match between the client's and therapist's perspective of the therapeutic event.

4.4.3 Significant events and outcome.

Timulak (2007) proposes that the significant events research paradigm assumes that there is a link between significant events and treatment outcome but that not many studies seem to look at this assumption and the ones that do find little correlation between the event and outcome. Furthermore, Timulak (2010) adds that significant events studies have limitations in that they often only summarise the impacts of the events and that those studies which do focus in-depth on events do so on a small number, often one at a time.

In a recent study of client-identified significant events, Timulak (2010) attempted to address the link between significant events and outcome by focusing on the issue of

whether ‘therapeutic change in a successful therapy case is visible in helpful significant events throughout the course of therapy’ (p.371). A total of 59 significant events from a 19-session course of therapy were analysed. The client was a young woman college student whose presenting issue was low mood. Therapeutic outcome was assessed by the Beck Depression Inventory (BDI; Beck, Steer & Garbin, 1988) and the Symptom Check List-90 (SCL-90) (Derogatis, Rickels & Rock, 1976). Results implied there were 14 mutually exclusive types of events. The most common types of event were those containing awareness or insight into explored problematic experiences (n=32) and those that characterized the client’s appreciation of therapy as a place to bring an important issue (n=6). Other frequent but less common types of event were those of where the client’s trust in the therapist played an active role (n=3), those where the client appreciated ‘exact understanding,’ therapist validation (n=3) and personal contact (n=3).

In a recent review, Timulak (2010) found that quantitative significant events and outcome studies were rare and found only one where there was a positive correlation between a significant events and outcome. The study was by Llewellyn (1988) and, included the participation of 40 clients who gave their views regarding helpful and unhelpful events after every therapy session. After therapy termination, clients reported on outcome. In total, 1,076 events were analysed from 399 sessions. The most frequently reported events for clients were those concerning reassurance/relief and problem solution. However, therapists reported the clients gaining of cognitive insight. After completion of therapy both clients and therapists reported the importance of personal contact.

4.4.4 Summary of Significant Events Research

The significant events research reviewed above would suggest that there is some evidence that some significant events are linked to outcome but evidence seems sketchy. In the studies that use recognized outcome measures, there was one participant. However, most of the studies find evidence that some significant events are helpful, with the most consistent theme being that of insight and self-understanding type events. Although not the most common type of event, personal contact and client-therapist type events appear to be important to therapy despite there being very little quantitative evidence for this.

4.5 Chapter Summary

In creating psychological measures the testing of reliability and validity is a complex issue with the concept of construct validity evidence being perhaps the most important aspect in designing any psychological test. Methodologies in testing for validity vary with the dominant approach still being traditional test methods (Classical Test Methods), although Item Response Theory and the Rasch Model are beginning to receive more attention recently. The Rasch Model is set apart from other quantitative psychological test theory in that it more closely mirrors physical measurement; it posits that the data must fit the model approach and therefore sets out to create an invariant, stable measure.

There appears to be a gap in significant events research in that there appear to be only a very small number of quantitative studies which use tried and tested outcome measures. The literature on web-based data collection varies widely in that results are not consistent; there is no consistent evidence to suggest that web-based data

collection has any issues that may cause validity issues (such as being able to generalize web-collected data to clinic (pencil and paper) collected data. In reviewing the different epistemologies, it would appear that the research philosophy underlying the current project is postpositivism. This project therefore uses methodology that best deals with the aims of the research in question.

This methodology literature review suggests that research into relational depth may greatly add to the psychotherapy research literature mainly because the topic of relational depth is still (at the time of writing) a little-researched topic. More specifically, the creation of a reliable and valid inventory designed to assess relational depth could provide a valuable research tool as it would enable quantitative research into relational depth and outcome. Furthermore, significant events research appears to lack outcome research quantitative studies. As relational depth can be seen to be a specific event, moment or experience, the concept conveniently lends itself to significant events research. An inventory that enquired about a significant event which then asked the respondent to reflect and quantitatively rate the event for presence of relational depth, would address many gaps in the research literature. It would have the benefit of being able to be administered with outcome measures where significant events in therapy and outcome could be researched. This would enable quantitative research on the relation between relational depth and outcome and help to explore the extent to which presence of relational depth in therapy contributes to positive therapeutic outcome. This would have the added benefit of quantitatively addressing significant events research.

Chapter 5: Study 1: The Prevalence and Characteristics of Relational Depth Events in Psychotherapy

5.1 Introduction to Chapter

The following chapter documents the first piece of research carried out as part of this whole research project and has recently been published in a research journal (see Wiggins, Elliott & Cooper, 2012). The current particular piece of research follows the definition of relational depth as an important event or moment, thus making it an example of significant event research. In light of the apparent gaps in research into relational depth, the main objectives of this piece of research were

1. to determine the extent to which moments of relational depth appeared to be present during significant therapy events
2. to investigate the reliability, validity and factor structure of the first questionnaire measure of relational depth
3. to test a new content analysis measure of relational depth in descriptions of significant events. This will be used as a means of assessing the validity of the measure and to explore some possible associated characteristics (gender, role and therapy duration) of relational depth events of which there is no known literature on these characteristics at the time of writing.

In this particular piece of research, two complementary measures of relational depth are introduced; one being a content analysis rating scale and the other the RDI. Using an internet-based survey of client and therapist accounts (n = 342), judges rated relational depth as present in over a third of significant therapy event descriptions.

Data from 80 of the participants had already been collected from the previous MSc study. A further 263 responded (as part of the PhD) in the current study making a total of 343.

Participants also completed the Relational Depth Inventory (RDI), for which reliability, validity and factor structure are reported. Relational depth events were more likely to occur in the presence of strong therapeutic alliance, and with female participants, but client or therapist role and therapy duration were not related to relational depth content or RDI. RDI items for connectedness, love, respect and intimacy were most strongly associated with relational depth content. This piece of research is documented in detail below.

5.2 Method

5.2.1 Participants

Eighty (23.3%) participants were male, 257 (74.9%), were female and 6 (1.2%) did not indicate gender. Of the 343 participants who took part in this study, 189 (55.1%) took part as therapists, 152 (44.3%) took part as clients and 2 (.6%) did not indicate. Of the therapists, 56 (29.6%) reported to be male, 131 (69.3%) reported to be female and 2 (1.1%) did not indicate. Of the clients, 24 (15.8%) reported to be male, 126 (82.9%) female and 2 (1.3%) did not indicate.

Ethnicity and socioeconomic status were not recorded. Note that due to the recruitment procedure (see Procedure below), it is likely that many of the clients in this study were actually therapists drawing on their experiences as clients [can this be specified]. It should also be noted that a precise estimate of response rate cannot be

calculated as it is not known how many people viewed the online questionnaire. However, due to the nature of the website software and its administration, it is known that 798 began the questionnaire and 343 completed and submitted their responses. It is also known that a total of 2,250 emails were sent (en masse) to various therapists listed on counselling and psychotherapy listings (such as the British Association for Counselling and Psychotherapy). Therefore, it can be said, as a very rough estimate, that the response rate was between 15% and 43%. This does not account for any individuals who may have viewed the survey website by chance and it does not account for those who may not have received emails due to spam filters.

5.2.2 Measures

Relational Depth Inventory (RDI). The RDI used in this study was earlier developed previously as part of an earlier unpublished study (Wiggins, 2007). Item creation involved raw data in the form of over 300 client and therapist descriptions of their experiences (meaning units) and definitions of moments of relational depth. These descriptions were subjected to grounded theory analysis where open-coding was used. Eventually four domains and two layers of categories emerged in each of the four domains. Questionnaire items were formed to represent the categories of the analysis. Thus, questionnaire items were designed to assess moments of relational depth. **(**See Chapter 1 for an in-depth documentation of this research and Table 5.1 in the current chapter for domains, categories and items.)

The RDI was presented as 'The Relationship Between Therapist and Client' and began with a question that asked potential respondents to describe, in their own

words, an important event they had experienced during a therapy session (see Appendix 1 for RDI). Then the respondent was asked to rate this significant event using a five-point scale (1: "not at all", 2: "slightly", 3: "somewhat", 4: "very much", 5: "completely") indicating the extent to which they experienced each of the specific qualities represented by the 64 questionnaire items. The questionnaire also asked for demographic data such as gender, role (whether client or therapist) and therapy duration overall (less than a month, 1-6 months, 6-12 months, 12 months – 2 years or over 2 years).

The RDI was designed to be different from other relational measures. Although the closest comparison is the Working Alliance Inventory (WAI), there are important differences between the two, both conceptually and operationally (focus on a specific relational moment vs. the overall relationship). There are many examples of these differences in each measure's items.

Where the WAI has items assessing liking, respect, appreciation and care, the RDI assesses stronger qualities of intimacy, love, mutuality and connection. In addition, where the WAI focuses on the tasks of therapy and the collaboration between client and therapist on goals of therapy, the RDI focuses on less tangible aspects of the therapeutic atmosphere such as spiritual and transcendent experiences.

Relational Depth Content Analysis. A rating manual was created that plainly set out instructions for rating significant event descriptions for presence of relational depth (see Appendix 2 for rating manual). Training consisted of reading the manual as

well as discussion between raters. The manual defined a relational depth event as, a moment, a series of moments, an experience, or feeling during which the participant appears to be describing "a state of profound contact and engagement between two people [between themselves and the other person], in which each person is fully real with the Other, and able to understand and value the Other's experiences at a high level". The scale was designed to be predominantly a clarity scale and not an intensity scale. Instructions in the manual asked raters to assign a score from 0 to 3 to the narrative descriptions of events, depending how clearly relational depth was deemed present (0: "clearly not present"; 1: "probably not present"; 2: "probably present"; 3: "clearly or strongly present").

Table 5.1: Questionnaire items showing order of presentation, and guiding categories and domains

No.*		**Guiding Category**	**Domain**
		Connected	Experience of Relationship
2.	We had things in common		
11.	Love		
12.	A meeting of minds		
24.	Both of us were connected in some way		
46.	Intimacy		
		Mutuality	
4.	There was give and take		
41.	Mutuality		
43.	I was at one with the other		
50.	Equality		
		Security	
14.	A sense of privacy in the relationship		
18.	Trusting the relationship		
19.	Safe		
		Heightened self	Experience of self
1.	Liberated		
2.	Flow		
10.	Spiritual		
20.	In an altered state		
23.	Magical		
36.	Expansive		
38.	Happy		
40.	I was transcendent		
48.	Spontaneous		
53.	Open		
64.	Enlightened		

5.	Intense feelings		
16.	Energised		
17.	Courageous		
21.	Revitalised		
22.	Scared	Invigorated self	
26.	Angry		
32.	Sexual		
56.	Exhilarated		
58.	Excited		
59.	Empowered		
13.	Opposing feelings at the same time		
25.	Weird feelings		
27.	Paradoxical	Immersed self	
45.	Immersed		
62.	Soulful (deep feelings and/or emotions)		
8.	A sense of being in the moment		
15.	Centred on the present		
31.	Self-value		
34.	Vulnerable	True self	
47.	In touch with self		
54.	I was being real		
63.	Aware of experience		
		Respect	Experience of/towards Other
6.	Other person respected me		
9.	Other person valued me		
49.	Value of other		
52.	Respect for other		
		Trust	
28.	Other person trusted me		
57.	Trust of other		
		Being available	
44.	I was there for the other		
61.	Other was available to me		
		Empathy	
38.	Empathy for the other person		
37.	Other was empathic towards me		
		Other being real	
42.	Other person was real		
		Dynamic	Experience of Atmosphere
60.	A transformative atmosphere		
		Peacefulness	
30.	A silent atmosphere		
39.	A still atmosphere		
51.	A timeless atmosphere		
7.	A meaningful atmosphere		
29.	An inexplicable atmosphere	Significance	
33.	An awesome atmosphere		
55.	A unique atmosphere		

No* Question number indicating order in which questionnaire item was presented.

The rating manual was designed in such a way that pairs of raters would rate descriptions. Raters were all trained therapists as well as being researchers. There were three raters in order that there was not the same two raters for all descriptions. Two raters were PhD students (including the first author) and one a university professor.

Further instructions were added for resolving discrepancies between raters. This was done in order that there would be one score per participant serving to facilitate statistical which meant the scores could be used as a variable. Here, the manual instructed raters in the case of 1-point discrepancies to average the final score (e.g. if one rater assigned a score of 1 and another a score of 2, the final score should be averaged to 1.5). If discrepancies were more than 1 (e.g. one rater assigned a 1 and the other a 3), raters discussed the case until consensus was reached. Ratings of the significant event descriptions resulted in Relational Depth Presence scores ranging from 0 to 3 (with increments of 0.5). This therefore can be seen to be a continuous variable and will be referred as *RD Presence* from here on.

Significant events were dichotomised at mean ratings of 1.5 (exactly halfway on the scale): One hundred and eighty-two events were rated as having a presence of relational depth greater than 1.5 (i.e., probably, clearly or strongly) and will be referred to as *RD Events* (relational depth events); 116 events were those rated at less 1.5 (probably not or clearly not) and will be referred to as *Non-RD events* (non-relational depth events). Twenty-three events resulted with halfway-point mean ratings of 1.5; in these cases, the mean ratings were retained for all analyses, with the

exception of the discriminant function analyses, where the raters met to decide whether to designate each event as either an RD Event (n= 5) or Non-RD Event (n = 18).

Working Alliance Short Form-Revised (WAI-SR). The revised short version of the Working Alliance Inventory (WAI-SR; Hatcher & Gillaspy, 2006) was added in order to explore the relationship between working alliance and those events characterised by relational depth. This version of the Working Alliance was the latest short version and consisted of 12 items. Each item utilised a 5-point Likert scale instead of the usual 7-point scale used in previous short versions. There were four items addressing Bond, four addressing Tasks and four addressing Goals. The WAI-SR items were presented after the 64 relational depth items, and only to clients.

5.2.3 Procedure

After ethics approval, a website was constructed using the survey software package Surveymonkey (2006). Links to this website were sent to colleagues of the researcher and to therapists listed on therapist directories including the British Association of Counselling and Psychotherapy website. In order for the online questionnaire to be strictly anonymous, neither email addresses nor Internet Protocol (IP) numbers, were collected by the researcher or the administrators. Potential respondents gave their informed consent by clicking on an 'I agree' option of the questionnaire item which asked whether they agreed (or not) to give their informed consent to continue. The RDI questionnaire asks respondents to focus on an important event that happened in a particular therapeutic relationship during a

therapy session, either with a client or therapist. This meant respondents could respond to the questionnaire from the perspective of a client (from their experience as a client) or from their perspective as a therapist. A question therefore asked 'In the particular relationship you have in mind, were you a therapist or client?' If a respondent clicked to indicate they were a therapist, the WAI-SR items would not be presented to them and would be skipped.

Responses were collected securely by the software administration package and automatically downloaded into a spreadsheet for later analysis using statistical software.

For the content analysis, after data had been collected, raters judged the significant event descriptions. Interrater reliability was then calculated for each of three sets of ratings, with a mean Cronbach alpha (equivalent to ICC (2,2); Shrout and Fliess, 1979) for the three sets of ratings of .79 (mean r or ICC(2,1) = .66).

5.3 Results

5.3.1 Relational Depth Significant Events

Out of a total of 144 descriptions of significant events reported by clients, 50 (34%) received RD Event Presence ratings of 2 or higher (i.e., relational depth probably or clearly present). Twelve event descriptions (8%) were rated as clearly present (a score of 3) and 38 (26%) rated as probably present. This suggests that relational depth is clearly present in 8% and probably present in 26% of what clients consider to be significant events in therapy. Examples of each of the four levels are given below:

Level 3 client examples (clearly or strongly)

Participant 150: *There was an interpersonal connection in the moment and my perceptual awareness changed.*

Participant 183: *My therapist cried with me.*

Participant 279: *For as long as I needed to, my counsellor was prepared to hold my hands. ... My realisation was that she was giving me something very special.*

Level 2 client examples

Participant 81: *A session where I was able to be myself as a small child and to gain the ability to comfort myself without feeling embarrassed by the process and where I felt emotionally held by the therapist.*

Participant 166: *In the past I had been very manipulative (because I had not felt very empowered). This was normalised by my counsellor and I felt less bad about myself with less shame and guilt.*

Level 1.5 (midpoint) client examples:

Participant 74: *the feeling of being understood*

Level 1 client examples

Participant 85: *This situation was a setting about the structure of my family. We did body therapy and I had to assemble the different persons and their relations to myself and each other.*

Participant 190: *remembering a traumatic event in my childhood.*

Participant 313: *confession of a long held secret*.

Level 0 client examples

Participant 79: *the therapist used a word that I felt it pathologised me and I felt shocked and offended*

Participant 80: *realisation of differences in moral values and politics between my therapist and me. I understood her to be far more conservative than I and began to experience her as possibly being judgmental. The therapeutic environment did not feel as safe as it had before.*

Participant 195: when the therapist told me what my problem was and wasn't prepared to hear my side of things, he wasn't prepared to hear what I thought my issue was.

Out of a total of 176 therapist descriptions of significant events, 67 (38%) were assigned a score or 2 or 3 indicating that relational depth is probably or clearly present in 38% of therapist significant events in therapy. Of these 24 (14%) were assigned a score of 3, indicating a clear or strong presence of relational depth, and 43 (24%) a score of 2, indicating relational depth was probably present. This indicates that relational depth is probably present in 24% and clearly present in 14% of therapists' significant events experienced during therapy sessions. As with the client descriptions, examples of each of the four levels are given below.

Level 3 therapist examples

Participant 17. *The relationship was such that in a pause of silence there was enormous depth.*

Participant 93. *... we felt very close to each other.*

Participant 94. *...we met beyond our roles.*

Participant 162. *We felt a deep connection in the silence.*

Level 2 therapist examples

Participant 142: *Being sexually abused as a child was an experience the client felt able to share and how it had impacted upon their relationship with their father.*

Participant 210: *For what felt the first time in a long time the client saw themselves differently, heard themselves affirming their own values*

Level 1.5 (midpoint)

Participant 153: Where a client wasn't feeling any empathy towards their children, but then in a particular session that all changed and she started to see her previous relationship through the eyes of her children.

Level 1 therapist examples

Participant 1: *The client shared the affects that the loss of his child had on his relationship with his partner.*

Participant 44: *She was nervous but was opened to an experiential exercise. My client was talking about her grandchild. I was very focused on it*

Participant 179*: I expressed how I long to reach out to the person behind the laughter and jokiness, but have difficulty with this.*

Level 0 therapist examples

Participant 31: *I did not feel I was connecting with this client at all.*

Participant 45: *The client became very angry with me because I would not give her the answer. Her experience of me was withholding.*

Participant 184*: I felt angry and needed to tell her she was playing with my beliefs.*

5.3.2 Relational Depth Inventory Scale Analyses.

Preliminary analyses. Items were retained or eliminated following standard procedures as outlined by (for example) Kline (1999). Pearson correlations were first run (using the data from the whole sample of clients and therapists) to test for redundancy (high correlations) among RDI items, in order to avoid inflation of internal consistency and artificial factors. Ten items correlated very highly (>.7) with multiple other items and were therefore omitted; theoretical relevance was also a factor in decisions to drop items. An example of a pair of highly intercorrelated items is, Item 7 ("an atmosphere that was meaningful") and Item 9 ("the other person valued me").

Next, because the RDI is conceptualized as measuring a single relationship experience, reliability analyses (Cronbach's alpha) were run to test for internal consistency on the 54 non-redundant items (this excluded the 10 redundant items which each had multiple correlations with other items) and to identify inconsistent items. Seven were found to reduce internal consistency: Items 13 ("opposing feelings at the same time"); 22 ("scared"); 25 ("weird feelings"); 26 ("angry"); 27 ("paradoxical'); 32 ("sexual"); and 34 ("vulnerable"). Cronbach's alpha rose from .95 to .97 when these items were dropped. It should be noted here that although these items did not contribute to the internal consistency of the RDI general factor variable, six of them (not including Item 32, "sexual") comprised Factor 4 (Scared/Vulnerable) and did yield a consistent subscale among themselves. This indicates that experiences associated with being scared and vulnerable may not be associated with experiences of relational depth or they may make up a different dimension of relational depth not consistent with the main factor here.

Factor analyses. In order to understand relational depth and its possible subfacets in the RDI, traditional exploratory principal axis factor analyses were conducted on the 54 non-redundant questionnaire items (KMO = .94; Bartlett's test, $p < .001$). The eigenvalue = 1 criterion was used initially, resulting in six factors; however, examination of the scree plot suggested a five factor solution. A varimax rotation was used to look for simple structure; the five-factor solution was fully interpretable and accounted for 53% of the overall variance. These results, including factor names and item loadings on their main factor, are shown in Table 5.2. Each factor was named appropriately according to the items (especially higher loading items) it comprised.

Factor 1 was named *Respect, Empathy & Connectedness* where the first two highest loading items referred to the informant's perception of their contribution to the relationship, e.g., Items 52 ("Respect for other") and 35 ("Empathy for the other person"). Other items included those that implied connectedness Item 46 ("Intimacy") and 24 ("Connected"). This was the largest factor with 15 items. The internal reliability of this factor was very high with an Alpha value of .94.

Factor 2 was named *Invigorated / Liberating* due to there being many items that implied this quality. The highest loading item was Item 21 (*Revitalised*), with other high loading items being Item 1 (*Liberated*), and Item 59 (*Empowered*). This factor had nine items and its internal reliability was very high with an Alpha value of .91.

Factor 3 was named *Transcendence* as most of its items concerned encounters that went beyond everyday experiences, for example, spiritual experiences. Item 29 (*Inexplicable Atmosphere*) and Item 51 (*Timeless Atmosphere*), were the highest loading items. Item 23 (*Magical*) and Item 43 (*I was transcendent*), were the next two highest loading items. This factor's internal reliability was very high with an Alpha value of .90.

Factor 4, was named *Scared / Vulnerable* due to the highest loading two items being Item 22 ("Scared") and Item 34 ("Vulnerable"). These six items yielded an internal of reliability of .79. It should be noted here that in a subsequent analysis, the items in this factor were found to be inconsistent with the rest of the scale (see Relational Depth Inventory Scale Analysis).

Factor 5, *Other person empathic/respectful*, was the smallest factor having only 3 items referring to the other person's contribution to the relationship: Item 37 ("Other empathic towards me") Item 6 ("Other respected me") and Item 61 ("Other was available to me"). This factor's internal reliability was high, with an Alpha value of .82.

In addition, factor analyses were carried out separately for clients and therapists, (see Tables 5.2 and 5.3). Each of these analyses resulted in three-factor solutions with many similarities and differences between each other..

Table 5.2. Results of varimax-rotated principal axis analysis on the 54 retained relational depth items factor

Items	1. Respect, Empathy & Connected ness	2. Invigorated /Liberating	3. Tran-scen-dence	4. Scared / Vulnerable	5. Other person empathic/ respectful
52.respect for other	**.78**	.20	.09	-.14	.24
35.empathy for the other person	**.73**	.13	.11	-.06	-.14
28. the other person trusted me	**.71**	.14	.12	-.05	.25
46.intimacy	**.67**	.13	.37	.09	.18
24.the both of us were connected in some way	**.67**	.18	.36	-.15	.15
50.equality	**.66**	.23	.15	-.21	.08
41.mutuality	**.64**	.23	.21	-.19	.25
43.I was at with the other	**.63**	.21	.40	-.19	-.02
54.I was being real	**.62**	.42	.04	-.01	.15
8. a sense being in the moment	**.59**	.31	.18	.04	.22
45.immersed	**.56**	.21	.32	.16	-.01
12.a meeting of minds	**.56**	.33	.19	-.19	.24
4.there was give and take	**.53**	.23	.08	-.17	.23
63.aware of experience	**.51**	**.42**	.17	.13	.05
15.centred on the present	**.50**	.31	.21	.02	.03
14.a sense of privacy in the relationship	**.46**	.02	.24	.16	.29
2.we had shared things in common	.39	.27	.13	-.31	.13
21.revitalised	.18	**.75**	.25	-.12	.15
1.liberated	.07	**.68**	.13	-.19	.33
59.empowered	.20	**.66**	.16	-.08	.37
56.exhilarated	.22	**.65**	.37	-.05	.06
16.energised	.38	**.64**	.16	-.09	-.05
64.enlightened	.35	**.61**	.27	.03	.14
60. a transformative atmosphere	.38	**.55**	.36	-.03	.17
38.happy	.15	**.54**	.23	-.35	.17
48.spont	**.44**	**.51**	.25	-.01	-.05
3. a sense of flow (a sense of smooth and continous change in myself)	.29	**.50**	.31	-.25	.14
47.in touch with self	**.47**	**.50**	.03	.04	.20
17.courageous	**.40**	**.49**	.08	.12	.11
31.self-value	**.39**	**.43**	.11	-.04	.38
29. an inexplicable atmosphere	.08	.08	**.70**	.28	.02
51. a timeless atmosphere	.36	.21	**.70**	-.09	.06
23.magical	.20	.43	**.68**	-.01	.02
40.I was transcendent (going beyond my ordinary limits)	.21	.40	**.61**	.14	.07
33.an awesome atmosphere	.27	.40	**.59**	.02	.01
39.a still atmosphere	.23	.06	**.59**	-.15	.28
30.a silent atmosphere	.05	.03	**.58**	.12	.08
20.in an altered state	.07	.26	**.56**	.18	-.09

11.love	.34	.21	**.55**	-.09	.34
10.spiritual	.27	.36	**.51**	-.11	.15
55. a unique atmosphere	**.42**	.22	**.49**	.13	.10
62.soulful	**.44**	.33	**.49**	.15	.22
36.expansive (unrestrained)	.34	**.46**	**.46**	.00	.07
32.sexual	.08	.05	.18	.05	.13
22.scared	-.11	-.12	.04	**.74**	.01
34.vulnerable	-.04	-.12	.16	**.69**	.08
13.opposing feelings at the same time	.01	-.09	-.07	**.64**	-.04
27.paradoxical (seemingly contradictory but nonetheless possibly true)	.04	.11	.17	**.57**	-.04
25.weird feelings	-.21	-.07	.41	**.54**	-.08
26.angry	-.36	-.10	-.13	**.52**	-.15
5.intense feelings	.22	.06	.26	.37	.04
37.the other was empathic towards me	.12	.23	.05	-.02	**.75**
6.the other person respected me	.41	.24	.10	-.10	**.64**
61.the other person was available to me	.31	.26	.26	-.07	**.53**
Variance explained (%)	17.06	13.82	11.85	6.25	5.05
Eigenvalue	9.21	6.93	6.40	3.37	2.73
Reliability[a]	.94	.91	.90	.79	.82

Note. Boldface indicates items with salient loadings $\geq$ .40 or ambiguously loading items.
[a]Cronbach's alpha calculated using items with a loading $\geq$ .40 and without ambiguities.

In each of the factor analyses for both clients and therapists (see Tables 5.3 and 5.4) there was a Respect/Empathy/Connectedness factor with Item 52 (*Respect for other*) as the highest loading item (Factor 1 for the client sample and Factor 2 for the therapist). In addition, this same 'Respect' factor in each case (for clients and therapists) also included items concerning the relationship including Item 46 (*intimacy*) Item 24 (*connected*) and Item 41 (*mutuality*). The main difference in this factor appeared to be that for therapists there were items assessing concern for other such as Item 35 (*Empathy for other*), while for clients there were items that reflected being cared for, such as Item 6 (*Other respected me*), Item 28 (*Other trusted me*) and Item 37 (*Other empathic towards me*).

Table 5.3. Results of varimax-rotated principal axis analysis on the 54 retained relational depth items for clients

	Factor		
Item	1. Respect	2. Invigorated / Transcendence	3. Weird / Scared
52.Respect for other	**.83**	.27	-.06
6.The other person respected me	**.82**	.25	-.10
28.The other person trusted me	**.80**	.12	.04
37.The other was empathic towards me	**.80**	.30	-.07
61.Other was available to me	**.80**	.25	-.06
41.Mutuality	**.72**	.28	-.06
46.Initmacy	**.72**	.16	.30
24.The both of us were connected in some way	**.69**	.31	.05
50.Equality	**.68**	.32	-.06
31.Self-value	**.67**	.45	.00
4.There was give and take	**.63**	.26	-.09
11.Love	**.63**	.31	.24
14.A sense of privacy in the relationship	**.63**	.12	.12
8.A sense of being in the moment	**.62**	.37	.11
12.A meeting of minds	**.62**	.38	-.01
62.Soulful (deep feelings and/or emotions	**.60**	.43	.31
43.At one with the other	**.60**	.36	.13
54.I was being real	**.59**	.45	.01
35.Empathy for other person	**.55**	.15	.13
47.In touch with self	**.53**	**.52**	.05
45.Immersed	**.53**	.13	.40
55.A unique atmosphere	**.52**	.27	**.44**
39.A still atmosphere	**.50**	.28	.17
63.Aware of experience	.47	.44	.19
26.Angry	-.45	-.32	.32
32.Sexual	.17	.12	.11
21.Revitalised	.26	**.86**	-.01
16.Energised	.28	**.79**	-.03
56.Exhilarated	.24	**.78**	.17
64.Enlightened	.41	**.71**	.03
1.Liberated	.41	**.70**	-.21
38.Happy	.29	**.70**	-.02
59.Empowered	.47	**.68**	-.12
3.A sense of flow (a sense of smooth and continuous change in myself)	.30	**.65**	-.08
60.transformative	.47	**.65**	.08
23.Magical	.22	**.62**	.43
48.Spontaneous	.36	**.59**	.18
10.Spiritual	.38	**.59**	.22
40.I was transcendent (going beyond my ordinary limits)	.30	**.56**	.46
33.An awesome atmosphere	.29	**.53**	.34
17.Courageous	**.42**	**.52**	.00
36.Expansive (unrestrained)	.44	.49	.30
51.A timeless atmosphere	.34	.47	.45
15.Centred on the present	.41	.45	.10

2.We had shared things in common	.39	.44	-.10
25.Weird feelings	-.25	-.09	**.70**
29.An inexplicable atmosphere	.10	.14	**.67**
22.Scared	-.05	-.36	**.60**
20.In an altered state	.01	.37	**.57**
5.Intense feelings	.12	-.03	**.51**
34.Vulnerable	.03	-.44	**.51**
13.Opposing feelings at the same time	-.08	-.15	.49
27.Paradoxical (seemingly contradictory but nonetheless possibly true)	.05	.14	.47
30.A silent atmosphere	.05	.16	.37
Variance explained (%)	24.32	19.76	8.56
Eigenvalue	13.13	10.67	4.62
Reliability[a]	.96	.95	.72

Note. Boldface indicates items with loadings > .50 as well as ambigous items.
[a]Cronbach's alpha calculated using items with a loading ≥ .50 and without ambiguities.

Table 5.4. Results of varimax-rotated principal axis analysis on the 54 retained relational depth items for therapists

	Factor		
Item	Transcendence / Invigorated	Respect	Scared / Vulnerable
23.Magical	**.78**	.25	.07
33.An awesome atmosphere	**.73**	.28	-.04
40.I was transcendent (going beyond my ordinary limits)	**.69**	.32	-.08
29.A inexplicable atmosphere	**.69**	.17	-.26
11.Love	**.66**	.28	.21
51.A timeless atmosphere	**.64**	.39	.13
30.A silent atmosphere	**.63**	.13	-.12
39.A still atmosphere	**.61**	.10	.14
20.In an altered state	**.61**	.11	-.10
56.Exhilarated	**.56**	.35	.19
36.Expansive (unrestrained)	**.56**	.44	.11
10.Spiritual	**.56**	.28	.13
62.Soulful (deep feelings and/or emotions)	**.56**	**.52**	-.13
3.A sense of flow (a sense of smooth and continuous change in myself)	**.55**	.30	.40
60.A transformative atmosphere	**.54**	**.45**	.06
55.A unique atmosphere	**.51**	.36	-.10
21.Revitalised	**.50**	.31	.26
64.Enlightened	.49	.46	.00
59.Empowered	.46	.36	.25
61.Other was available to me	.46	.36	.19
1.Liberated	.46	.20	.36
32.Sexual	.23	.03	-.04
52.Respect for other	.03	**.72**	.17
35.Empathy for the other person	.04	**.71**	.08
54.I was being real	.10	**.66**	.10
46.Intimacy	.35	**.65**	-.03

24. The both of use were connected in some way.	.37	**.62**	.12
41.Mutuality	.25	**.61**	.24
8. A sense of being in moment	.24	**.61**	-.02
28.The other person trusted me	.15	**.60**	.10
45.Immersed	.33	**.58**	-.06
15.Centred on the present	.19	**.57**	-.05
43.At one with the other	.39	**.55**	.17
12.A meeting of minds	.31	**.53**	.41
50.Equality	.15	**.52**	.23
5.Intense feelings	.20	**.52**	-.25
47.In touch with self	.11	**.50**	.11
63.Aware of experience	.29	.49	-.06
17.Courageous	.25	.44	-.12
16.Energised	.32	.42	.06
48.Spontaneous	.36	.42	.10
14.A sense of privacy in the relationship	.21	.42	-.27
6.The other person respected me	.25	.41	.31
4.There was give and take	.18	.37	.31
31.Self-value	.25	.36	.20
37.The other was empathic towards me	.20	.32	.30
22.Scared	.08	-.04	**-.66**
27.Paradoxical (seemingly contradictory but nonetheless possibly true)	.21	.02	**-.64**
34.Vulnerable	.31	.19	**-.62**
13.Opposing feelings at the same time	-.12	.01	**-.60**
38.Happy	.37	.10	**.58**
25.Weird feelings	.43	-.04	**-.55**
26.Angry	-.08	-.08	-.46
2.We had shared things in common	.15	.34	.46
Variance explained (%)	17.40	19.20	4.25
Eigenvalue	9.39	9.29	4.26
Reliability[a]	.92	.90	.59

Note. Boldface indicates items with loadings > .50 as well as ambigous items.
[a]Cronbach's alpha calculated using items with a loading $\geq$ .50 and without ambiguities.

On inspection of Figures 5.3 and 5.4, in addition to a respect factor for both clients and therapists, another factor in each analysis included items concerning experiences of transcendence, such as Item 23 (“Magical”) and Item 10 (“Spiritual”). However, for clients these items were much lower down the factor with many items assessing invigorating experiences much higher. For therapists, the reverse was true where items assessing invigorating experiences were much lower than those assessing

transcendence. Therefore for clients this factor was named "Invigorated /Transcendence" and for Therapists "Transcendence/Invigorated."

5.3.3 Profiling significant events with a presence of relational depth.

In order to produce a profile of relational depth events, a discriminant function analysis was conducted using the 54 retained RDI items as predictors of group membership (this included the seven items that were dropped from the main scale in the internal consistency analyses). The groups in question were cases that were earlier rated, in the content analysis, as RD Events and Non-RD Events. The discriminant function significantly differentiated between RD Events and Non-RD Events (Multiple R = .44, Wilk's Lambda = .66, df = 54, p<.00001), indicating that the questionnaire items, as a whole, successfully predicted presence vs. absence of relational depth in significant event descriptions. The discriminant function successfully predicted group membership for 83.2% of cases (80.50% accuracy for RD events and 84.90% for non-RD events). Interpretation of the function was made using Pearson correlation coefficients between items' and RD Presence ratings, with RDI. Items were considered to be significantly correlated with RD Presence if their Pearson correlation coefficient was at least moderate in size (.30 or above; Cohen, 1988). A total of 26 items met this criteria. It should be noted that for each participant the 26 RDI items were totalled and mean average calculated for use in subsequent analyses. This therefore constitutes another variable and will be referred to as RDI Index from here on.

Frequencies were also calculated for RDI Index score; 11% of respondents scored below a 2, 20% scored from 2 but less than 3, 46% 3 but less than 4, and 23% were 4 to 5. This indicates that 23% of respondents experienced relational depth either 'very much' or 'completely' and 46% experienced it 'somewhat' to 'very much' indicating a higher rate than the content ratings would suggest. Table 5.5 contains various statistics for every questionnaire item. The first column gives Pearson correlation coefficients for correlations between each item's score and its RD Presence rating (RD Presence ratings were content ratings of significant event descriptions). The second column gives correlation coefficients between item score and RDI Index score. The next four columns give the item means and SDs for those descriptions classified as RD Events vs non-RD Events (RD Events and RD non-Events refers to the dichotomized cases where a rating of less than 1.5 was classed as a non-RD Event and more than 1.5 an RD Event). The last column shows effect sizes for the difference between RD vs. non-RD Events.

Table 5.5 Results of discriminant function analysis showing item correlations, means, standard deviations and effect sizes

	Corr RD Pres	**Corr RDI index**	**RD Event mean**	**RD Event SD**	**Non-RD Event mean**	**Non-RD Event SD**	**Effect size**
24. The both of us were connected in some way	0.47	0.79	4.03	0.95	3.08	1.15	0.90
11. Love	0.46	0.71	3.60	1.31	2.38	1.34	0.92
52. Respect for other	0.40	0.75	4.36	0.78	3.78	1.16	0.60
46. Intimacy	0.40	0.75	3.89	1.12	2.93	1.22	0.82
6. The other person respected me	0.40	0.66	4.41	0.74	3.81	1.14	0.64
41. Mutuality	0.39	0.75	3.84	1.03	3.08	1.21	0.68
61. Other was available to me	0.39	0.67	3.92	1.23	3.25	1.35	0.52
43. I was at one with other	0.38	0.78	3.39	1.39	2.64	1.38	0.54
8. A sense of being in the moment	0.37	0.69	4.59	0.64	3.99	1.13	0.67
12. A meeting of minds	0.37	0.73	3.65	1.17	2.97	1.17	0.58

39. A still atmosphere	0.37	0.55	3.02	1.40	2.08	1.07	0.76
28. The other person trusted me	0.36	0.71	4.23	0.85	3.78	1.16	0.52
60. A transformative atmosphere	0.36	0.73	3.80	1.19	3.20	1.29	0.48
51. A timeless atmosphere	0.36	0.74	3.22	1.43	2.33	1.36	0.64
23. Magical	0.35	0.67	2.61	1.42	1.94	1.25	0.49
62. Soulful (deep feelings or emotions)	0.35	0.75	3.84	1.16	3.22	1.35	0.49
55. Unique atmosphere	0.34	0.64	3.95	1.10	3.13	1.30	0.68
50. Equality	0.33	0.72	3.70	1.21	3.12	1.30	0.46
3. a sense of flow (a sense of smooth and continuous change in myself)	0.33	0.64	3.28	1.26	2.62	1.34	0.51
31. Self value	0.33	0.68	3.73	0.98	3.30	1.14	0.41
33. An awesome atmosphere	0.32	0.64	2.81	1.49	2.13	1.29	0.49
48. Spontaneous	0.32	0.67	3.75	1.16	3.10	1.22	0.55
54. I was being real	0.31	0.66	4.47	0.82	4.03	1.04	0.47
45. Immersed	0.31	0.63	3.66	1.29	2.99	1.37	0.50
10. Spiritual	0.31	0.67	3.20	1.52	2.50	1.37	0.48
4. There was give and take	0.30	0.63	3.58	1.08	2.86	1.27	0.61
21. Revitalised	0.29	0.64	3.07	1.16	2.72	1.30	0.28
1. Liberated	0.29	0.49	3.35	1.24	2.81	1.35	0.42
40. I was transcendent (going beyond my ordinary limits)	0.29	0.63	2.97	1.50	2.31	1.31	0.47
37. The other empathic towards me	0.29	0.47	3.59	1.36	2.98	1.42	0.44
2. We had shared things in common	0.28	0.50	3.09	1.39	2.32	1.29	0.57
36. Expansive (unrestrained)	0.28	0.65	3.30	1.35	2.42	1.31	0.66
15. Centred on the present	0.27	0.58	4.14	0.95	3.63	1.22	0.47
47. In touch with self	0.27	0.60	4.03	0.90	3.73	1.03	0.29
64. Enlightened	0.26	0.68	3.55	1.23	3.19	1.36	0.28
63. Aware of experience	0.26	0.58	4.24	0.82	3.90	1.06	0.36
59. Empowered	0.25	0.61	3.41	1.21	3.08	1.30	0.26
30. A silent atmosphere	0.25	0.37	2.53	1.52	1.97	1.09	0.43
35. Empathy for the other person	0.24	0.59	3.85	1.37	3.41	1.46	0.31
16. Energised	0.23	0.60	3.74	1.15	3.28	1.34	0.37
56. Exhilarated	0.21	0.66	2.89	1.32	2.58	1.26	0.24
38. Happy	0.19	0.49	2.89	1.39	2.58	1.28	0.23
29. An inexplicable atmosphere	0.17	0.43	2.91	1.49	2.33	1.36	0.42
20. In an altered state	0.17	0.38	2.80	1.50	2.53	1.28	0.14
5. Intense feelings	0.16	0.33	4.30	0.71	4.03	0.98	0.23
14. A sense of privacy in the relationship	0.15	0.52	3.90	1.10	3.43	1.26	0.40
17. Courageous	0.10	0.56	3.50	1.17	3.40	1.31	0.08
32. Sexual	0.08	0.15	1.33	0.79	1.29	0.76	0.05
34. Vulnerable	-0.03	-0.05	2.87	1.38	2.80	1.42	0.05
27. Paradoxical (seemingly contradictory but nonetheless possibly true)	-0.05	0.13	1.89	1.27	1.90	1.15	0.00
25. Weird feelings	-0.17	-0.04	1.78	1.19	2.05	1.24	0.22
22. Scared	-0.20	-0.20	2.02	1.15	2.33	1.28	0.17
13. Opposing feelings at the same time	-0.21	-0.10	2.19	1.39	2.56	1.41	0.26
26. Angry	-0.37	-0.40	1.41	1.41	1.93	1.31	0.47

Note Italicized items were dropped from the revised instrument

The two items that most strongly discriminated between RD Events and Non-RD Events were Item 24 ("Both of us were connected in some way"), and Item 11 ("Love"). The Pearson coefficients for the correlation between these items and RD Presence were .47 and .46 respectively. The effect size (Cohen's d) for the difference between RD and non-RD Events was .90 and .92 respectively for these two items. After these two items, the next most discriminating items were Item 46 ("Intimacy") and Item 39 ("A still atmosphere") with effect sizes being .82 and .76 respectively. The Pearson correlation coefficient for these items was .40 and .37. One item, "Angry" (Item 26), yielded a negative correlation of -.37 with RD Presence, indicating that an absence of anger was associated with the global impression of relationship depth**.**

*Convergent validity: Working Alliance (WAI-SR) and RDI Index***.** In order to explore criterion-related evidence for of the RDI**,** Pearson Correlation Coefficients were calculated in order to assess the extent to which the RDI Index (i.e., the mean score across the 26 retained RDI items) was associated with both RD Presence (i.e., content ratings of significant events for relational depth presence) and the WAI-SR. It should be noted, that as therapists did not complete the WAI-SR, correlations involving the WAI-SR were performed using the client sample only. For RD Presence and WAI–SR a medium correlation was found ($r = .33, N=142, p < .01$) indicating that relational depth events are, to a moderate degree, associated with the quality of the working alliance, with. In other words, this indicates that the stronger the working alliance is, the stronger relational depth event is. There was a moderate to large correlation between RD Presence and RDI Index ($r = .50, N=320, p < .01$)

suggesting that the RDI Index is assessing relational depth in significant events to a considerable degree. For RDI Index and WAI-SR, a large correlation was found (*r =.72, N=150, p <.01)* indicating that relational depth events, as assessed by the RDI items, are more likely to occur where there is strong therapeutic alliance.

5.3.4 Factors influencing the presence of relational depth in significant events

Informant Role: Client vs Therapist. In order to investigate differences in RD Presence and RDI Index between clients and therapists, independent samples *t*-tests were used. The difference between RD Presence for therapists (mean = 1.40, SD = .92, N = 176) and clients (mean = 1.23, SD = .97, N=144) was small and not significant (t = 1.604, df = 318, p =. 11; d =.18). This result indicates relational depth was no more or less likely to occur for clients than therapists during significant events in therapy. For RDI Index, however, therapists' scores (mean = 3.48, SD = .75, N = 187) were higher than clients' (mean = 3.01, SD = 1.03, N = 152) and the difference was medium-sized and significant (t = 4.71, df = 270, p <.001; d = .52). This suggests that the RDI Index as a whole is more sensitive to differences between clients and therapists than RD Presence ratings.

In order to understand the factors and role more clearly, T-tests were conducted to investigate differences between clients and therapists on each of the resultant five factors from the factor analysis. For differences to be significant, p values would have to be equal to, or less than, the Bonferroni adjusted significant level of .01. Results are shown in Table 5.6.

Table 5.6. Client and therapist differences on RDI factors

Factor name	Therapists Mean	Therapists SD	Clients Mean	Clients SD	t-values	Effect size
1.Respect, Empathy & Connectedness	3.84	0.70	3.16	1.04	6.94	.78*
2. Invigorating / Liberating	3.14	0.88	2.87	1.17	2.35	.26
3. Transcendence	2.68	1.05	2.41	1.02	2.44	.31
4. Scared / Vulnerable	2.04	0.88	2.41	0.93	-3.74	.44*
5. Other Person Empathic / Respectful	3.24	0.92	3.78	1.32	-4.21	.48*

Note. *p < .05

As can be seen, there were significant differences between clients and therapists on three of the five factors. These indicate that therapists, compared to clients, were more likely to experience respect, empathy and connectedness (Factor 1) during significant events in therapy. It also suggests that clients, compared to therapists, were more likely to experience being scared and vulnerable (Factor 4) and more likely to experience the other person as being empathic and respectful toward them during significant events.

Informant Gender: Males vs Females. The mean RD Presence rating for male respondents was 1.13 (SD = .85, N = 76) and for females it was 1.38 (SD = .97, N = 240). Although the effect size was very small (d = .14), females' descriptions were rated significantly higher than males' (t = -2.17, df = 141.03, p = .03), indicating a slight gender effect in favour of females**.** For RDI Index the mean for females was 3.30 (SD = .95, N = 256) and for males it was 3.18 (SD = .80, N = 79), but this difference was not statistically significant (t = -1.04*,* df = 333*,* p = .30, d = .13*)*.

Role x Gender Interaction. In order to investigate whether there was a gender x role interaction on RD Presence or RDI Index, two 2 (role: client vs therapist) x 2 (gender: male vs female) univariate ANOVAs were conducted. For RD Presence, a significant interaction was not found [F (1, 312) = .01, p > .05] suggesting that role and gender do not interact significantly to affect whether relational depth is present during significant events in therapy. For RDI Index score, a significant interaction was not found [F (1,331) = .19, p > .05] indicating that role and gender do not interact in relational depth experiences assessed by the RDI Index. Please see Table 5.7 for means and standard deviations. Effect sizes for gender comparisons ranged from a *d* of .16 to .35.

Table 5.7. Role x Gender Effects on Relational Depth Variables

		Gender					
		Male (N=76)		Female (N=240)		Overall (N=316)	
Variable	Role	Mean	SD	Mean	SD	Mean	SD
RD Presence	Therapist (N=174)	1.20	0.88	1.48	0.93	1.40	0.92
	Client (N=142)	0.96	0.78	1.27	1.00	1.22	0.97
	Overall mean	1.13	0.85	1.38	0.97	1.32	0.95
RDI Index	Therapist (N=174)	3.30	0.76	3.56	0.75	3.48	0.76
	Client (N=142)	2.89	0.82	3.04	1.06	3.01	1.02
	Overall mean	3.18	0.80	3.30	0.95	3.28	0.92

Therapy duration. It should be noted that therapy duration referred to overall length of therapy and not necessarily how far into therapy the significant event happened. A between-subjects one-way ANOVA was used to look at whether Therapy Duration (divided into five levels) was associated with ratings of relational depth in significant event descriptions (RD Presence); there was no significant difference (F (4,310) = 1.55, p = .19, partial eta^2 = .02). A Pearson correlation was also performed and the correlation between RD Presence and Therapy Duration was also not significant (r =

-.03, $N = 321$, $p = .649$). This indicates that presence of relational depth is not influenced by overall duration of therapy (i.e., longer did not mean deeper events).

5.4 Chapter Summary

Study 1 provided evidence about the nature of relational depth. Unlike previous studies, experiences of love appeared to characterise relational depth. This first study also helped to considerably revise the RDI into a much shorter and more accessibly worded inventory; the RDI-Revised (RDI-R). This first study provides rich evidence about the nature of relational depth on a larger scale than previously, using two new quantitative measures of this conceptualization of therapeutic relatedness, each of which can be used either as a continuous variable or index for studying associations with other variables (such as therapeutic alliance or gender) or as a dichotomy, for identifying significant events marked by the presence of relational depth. Also the likely sub-factors of relational depth were laid out and established a strong but not identical relationship between therapeutic alliance and relational depth. Gender was seen to have a limited roles as a determinant of relational depth events.

The results of Study 1 are further discussed in the main discussion (Chapter 9).

Chapter 6: Revising the Relational Depth Inventory

6.1 Introduction to Chapter

This section of the project is intended to document exactly how the Study 1 data collected from the 64–item RDI was used to create the 24-item Relational Depth Inventory-Revised (RDI-R). The new version of the RDI, instead of being created as a general therapist or client measure (where 'other' was referred to), was instead intended to become a client-only version. Therefore instead of the word *other*, the word *therapist* was used in the wording of the items when the self was referring to the other member of the therapy dyad. In other words the self would always be the client and the other would always be the therapist. This chapter, therefore, documents in detail, how the 64-item RDI became the 24-item RDI.

6.2 Using Study 1

A combination of quantitative and qualitative methods were used in order to reduce the 64 items to 24. Consensus of opinion was also used in deciding how to reword some items. However, the results form Study 1 were drawn heavily upon and largely guided the creation of the 24-item version RDI.

Firstly, in Study 1, 26 items were found to have significant correlations (>.3) with researchers' ratings of relational depth presence (referred to as RD Presence from here on) in the client described helpful events (see Table 6.1 below). Consequently, it was intended these would be used in any future RDI. Secondly, 10 items ('revitalised', 'liberated', 'transcendent', and 'other empathic towards me') did not

make the .30 cut-off point with from .25 to .29. However, these items were considered to be theoretically relevant to the concept of relational depth and were therefore included in the creation of the shorter version. In other words strict adherence to only keeping items with a correlation of .3 or more seemed impractical. Thirdly, another consideration to make when creating new items was wording. This became apparent after some feedback at the Strathclyde Research Clinic when administrators reported that clients did not understand some of the phrases as many thought the wording was 'therapist jargon'. This was not identified earlier as the preliminary 64-item version had been completed by a sample of participants, a large proportion of whom were most probably therapists who would have understood and been more comfortable with the terminology far more readily.

Table 6.1. Items from 64-item version and their correlations with RD presence

Item	**PC***	**Item**	**PC***
24. Both of us connected	.47	3. Flow	.33
11. Love	.46	31. Self-value	.33
52. Respect for other	.40	33. Awesome atmosphere	.32
46. Intimacy	.40	48. Spontaneous	.32
6. Other respected me	.40	54. I was being real	.31
41. Mutuality	.39	45. Immersed	.31
61. Other was available to me	.39	10. Spiritual	.31
43. At one with the other	.38	4. Give and take	.30
8. Being in the moment	.37	21. Revitalized	.29
12. Meeting of minds	.37	1. Liberated	.29
39. A still atmosphere	.37	40. Transcendent	.29
28. Other trusted me	.36	37. Other empathic towards me	.29
60. Transformative atmosphere	.36	2. We had shared things in common	.28
51. Timeless atmosphere	.36	36. Expansive (unrestrained)	.28
23. Magical	.35	15. Centred on the present	.27
62. Soulful	.35	47. In touch with self	.26
55. Unique atmosphere	.34	64. Enlightened	.26
50. Equality	.33	59. Empowered	.25

Note. PC = Pearson Correlation Coefficient with Relational Depth Presence.

The feedback exercise was a very simple one which simply used the 36 items that most correlated with RD Presence as an initial shorter version. These are shown in Table 6.1 along with other lower correlating items. The rewording of items was conducted by using feedback from fellow researchers as well as using consensus of opinion. To begin creating a shorter version, the 36 items which most correlated with Relational Depth Presence were presented in the same format as the 64-item version (see Appendix for 64-item version). This was administered to therapists and colleagues of the researcher, some of whom were not therapists. Administrators at the Strathclyde Research Clinic were also given this 36-item version. Various comments and feedback were noted and items were either reworded or deleted. Some new items were also added due to the feedback. This is documented in the next section.

6.3 Using Feedback

To clarify exactly how items were reworded or deleted or added, Table 6.2 is shown below showing old and new versions of items as well as showing items added or deleted. A brief explanation of reasons for change, addition or deletion, are included in the same table (Table 6.2) with a more comprehensive explanation below. In the revision of the RDI, priority was given to making the set of items as short as possible whilst also being comprehensive in that all theoretical aspects of relational depth could be included. In addition, it was important that items were able to be understood by most people as this new inventory would be administered to participants from all educational backgrounds and therefore accessibility and clarity were of upmost importance. Although Table 6.2 documents to some degree the addition, deletion and rewording of items, for extra clarification I explain in detail

below how and why items were treated as they were. However, I do not document the simple rewording of items as I hope these are self- explanatory in the table (Table 6.2). Instead I have documented the major additions and deletions.

Table 6.2 Rewordings, deletions and additions of RDI

Old Item **The stem 'I felt'…was included in the instructions in the old version**	**New Item** **The stem 'I felt' was used as part of the item in most items as it seemed clearer this way.**
24. Both of us were connected in some way	10. My therapist and I were both connected in some way
11. Love	24. I felt a warm personal bond as fellow human beings
52. Respect for other	19. I felt respect for my therapist
46. Intimacy	16. My therapist and I were close to each other
6. Other respected me	3. I felt my therapist respected me
41. Mutuality	Deleted due to many pairs of items assessing perceived mutuality (e.g. Items 19 and 3 above).
61. Other was available to me	22. I felt my therapist was there for me.
43. I was at one with the other	Deleted as the reworded intimacy item, Item 16 ('close to each other') was very similar and therefore was thought to be redundant
8. Being in the moment	4. I felt I was living in the moment
12. Meeting of minds	7. My therapist and I both knew what was in each other's mind
39. A still atmosphere	Deleted due to feedback – 'weird'
28. Other person trusted me	11. I felt my therapist trusted me
60. Transformative atmosphere	Deleted – thought this may be too linked to outcome
51. Timeless atmosphere	18. I felt as if time had stopped
23. Magical	9. I felt a kind of magic happened
62. Soulful	Deleted due to feedback where people thought it was weird, strange.
55. Unique atmosphere	Deleted
50. Equality	17. I felt my therapist and I were equal.
3. Flow	Deleted
31. Self-value	Deleted
33. Awesome atmosphere	13. I felt the atmosphere was kind of awesome
48. Spontaneous	Deleted
54. I was being real	20. I felt I was being genuine with my therapist
45. Immersed	Deleted
10. Spiritual	6. I felt in touch with my spiritual side
4. Give and take	2. There was give and take between my therapist and myself
21. Revitalised	8. I felt more alive
1. Liberated	1. I felt a sense of freedom
40. Transcendent	15. I felt I was going beyond my ordinary limits
37. Other empathic towards me	5. My therapist understood what it was like for me
59. Empowered	21. I felt a sense of having my own power.
64. Enlightened	23. I felt I had better understanding of myself and others
No old item – added in order to assess perceived mutuality of Item 20.	12. I felt my therapist was being genuine with me
No old item – added in order to assess perceived mutuality of Item 5.	14. I felt I understood what it as like for my therapist

Item 41: 'Mutuality'

A major change was the deletion of the mutuality item, Item 41 (Mutuality). This was dropped because, after some additions, there is effectively a perceived mutuality scale incorporated into the RDI. These are Items 19 and 3 which is designed perceived mutual respect, Items 20 and 12 mutual genuineness and Item 5 and 14, mutual empathy. The additions here (12 and 14) are further documented below.

Item 43: 'I was at one with the Other.'

The next major change was another deletion, that of Item 43 ('I was at one with the other'). This was done because it was thought that the newly reworded item which assessed intimacy Item 16 ('My therapist and I were close to each other') also now assessed an 'at oneness'. Furthermore, the concept of being 'at one' with somebody may have romantic connotations.

Item 39 ('A still atmosphere)

Item 39 ('A still Atmosphere) was dropped due to some feedback saying it was 'weird'.

Item 60 ('Transformative atmosphere')

Item 60 (Transformative Atmosphere) was deleted as after some discussion with other researchers this was felt to be assessing the process of therapy rather than the relational qualities between therapist and client.

Item 62 ('Soulful')

Item 62 (Soulful) was deleted due to feedback where again people thought that it was ambiguous. Despite being presented with the words 'deep feelings and/or emotions' this item was considered to be ambiguous from due to feedback.

Item 55 ('Unique Atmosphere')
Item 55 (Unique atmosphere) was deleted due to people saying they thought this item was ambiguous in that they thought the wording was unusual.

Item 3 ('Flow')
Item 3 ('Flow') was deleted due to feedback as some thought it was therapist jargon and that many clients out there in the 'real world' would not know what this meant or would be put off. This was despite having the words 'a sense of smooth and continuous change in myself.'

Item 31. ('Self-value')
Item 32 (Self-value) was deleted as this was felt by many to have very little to do with relational depth and more to do with self-experiences.

Item 48. ('Spontaneous')
The spontaneous item was deleted due again to feedback. many said they thought some clients would not know what this meant or that they may be put off by such complex words.

Item 45. ('Immersed')
Item 45 ('Immersed') was thought to be an ambiguous item and many thought that they did not know what it meant. Many also said they thought that clients would be put off by

Item 2. ('We had shared things in common')

Item 2 (‘We had shared things in common’) yielded a low correlation (.28) with RD presence and was thought be too much like the reworded ‘connection’ Item 10 (‘My therapist and I were both connected in some way’) so.

Item 36. Expansive

Item 36 was not used for two reasons, firstly because it yielded a low correlation with Relational Depth Presence and because the item appeared ambiguous. It was therefore left out.

Item 15. Centred on the Present

Item 15 was left out of the revised version due its low correlation with RD Presence and due to the fact that it was very similar to the reworded Item 4 (‘I felt I was living in the moment’).

Item 47. In touch with self

Item 47 was deleted because of its low correlation with RD Presence and because, according to feedback, it has little to do with relational depth.

Retained despite being < .3 with RD Presence

Item 64 (‘Empowered’) became Items 21 (‘I felt a sense of having my own power). Item 21 was designed to assess the experience of being empowered and was reworded. This item in the original questionnaire yielded a low correlation with RD presence (.26). However, feedback suggested it was relevant to relational depth as many people said they liked it and that it should be kept in.

Item 59 (‘Enlightened’) became Item 23 (‘I felt I had a better understanding of myself and others’). This item designed to assess the experience of feeling enlightened. As with the above item, this was retained because feedback suggested it was relevant to relational depth and that people would understand the item.

Additions on revised 24-item RDI.

Item 12 (‘I felt my therapist was being genuine with me’). This item was added in order that perceived mutuality may be assessed for genuineness; so that it would match Item 20 (I’ felt I was being genuine with my therapist’)

Item 14 (‘I felt I understood what it was like for my therapist’). Item 14 was added for the same reason as Item 12 above; so that perceived mutuality could be assessed for empathy.

6.4 Chapter Summary

After all the above rewording, amendments, deletions and additions the final 24 items were used to create the 24-item inventory. As with the 64-item version, wording at the beginning of the questionnaire asked the respondent to think of a helpful event in therapy then to rate it using the 24 items. This revised and shorter version of the inventory, known as the Relational Depth Inventory-Revised (RDI-R), was now ready to be used in a validity study and in an outcome study. Two such studies were carried out and are described in the following chapters.

Chapter 7: Study 2: Evidencing and Improving Reliability and Validity using the Rasch Model

7.1 Introduction to Chapter

Study 2 (Chapter 7) is essentially composed of Rasch analyses designed to test the reliability and validity of the Relational Depth Inventory-Revised (RDI-R). As already explained in the methodological review (Chapter 4), the philosophy of this research project is that any psychological measure should have an invariance quality. Therefore, it is my view that the Rasch Model is the most straightforward, direct model to use when exploring research which attempts to explore reliability and validity evidence. The various Rasch statistics described in the methodological literature review (Chapter 4), were carried out for this particular study. Data were Rasch analysed and the integrity of the response scale and potential item reductions are explored.

In order to address the reliability and validity of the RDI, this study attempts to answer the following questions.

1. Does the 5-point rating scale function as intended?
2. How many distinct groups of clients are there experiencing different levels of relational depth?
3. Do all 24 items form a single unidimensional measure?
4. Does the item hierarchy make sense theoretically?
5. Does the RDI-R target persons appropriately; or are there persons it does not target?
6. Does the WAI target persons in a similar or different way to the RDI? Specifically, do RDI items map at generally higher levels of the measure than the WAI.

7.2 Method

7.2.1 Participants

Participants consisted of 163 bona fide (non-therapist) clients. The number of female clients was 134 (82%), male clients, 25 (15%) and 4 (3%) did not specify gender. Age range was between 14 years and 68 years with the average age being 38. All participants were clients who were not therapist-clients, in other words they were neither trained nor practising therapists. Of the 163, 120 (74%) were from an online sample (see below) and 43 (26%) were from the Strathclyde Research Clinic. As the online sample was completely anonymous, it is not known from what geographical location they were from. However, the Strathclyde Research Clinic sample were clients drawn from the Glasgow area of Scotland, UK. The 43 clients from the research clinic included 27 Practice-Based Clients (PB clients) and 16 Social Anxiety Clients.

For one analysis (the WAI and RDI analysis), to answer Question 5, only data from the 43 clients from the research clinic were used. This is because the online sample did not complete any WAIs. In addition, in order to increase case numbers and therefore statistical power, non-independent cases were used and so there was the following.

1 x data set for each of 25 clients = 26 cases
2 x data sets for each of 8 clients = 16 cases
3 x data sets for each of 5 clients = 15 cases
4 x data sets for 1 client = 4 cases
Total 61 cases

7.2.2 Procedures.

After ethics approval, by the University Ethics Committee and by The NHS Ethics Service, a website was constructed using the survey software package

Surveymonkey. In order to recruit bona fide clients and to avoid therapist-clients, links to this website were sent to self-help websites such as Psychlinks Psychology Self-Help & Mental Health Support Forum, Uncommon Knowledge, The Forum Site, HealthyPlace.com and Healthypages. As well as this, a question asked participants whether they were therapists. In order for the online questionnaire to be strictly anonymous, the researcher and the administrators collected neither email addresses nor Internet Protocol (IP) numbers. Potential respondents gave their informed consent by clicking on an 'I agree' option of the questionnaire item which asked whether they agreed (or not) to give their informed consent to continue.

7.2.3 Measures

The Relational Depth Inventory - Revised

The Relational Depth Inventory - Revised (RDI-R) was used for the purposes of this study. How the RDI (64-item inventory) became revised is explained in Chapter 6. The RDI-R consists of a question that asks participants to describe a helpful event in therapy and then asks them to rate that event using 24 items. Each of the 24 items was presented with a Likert scale with scores ranging from 1 to 5 (1 = not at all, 2 = slightly, 3 = somewhat, 4 = very much, 5 = completely). Examples of items include 'I felt my therapist respected me' and 'I felt my therapist and I were both connected in some way.' Please see Appendix 3 for the complete RDI-R.

7.3 Results

For all the analyses below, the Rasch rating scale model was used as the method of data analysis using the software WINSTEPS (Linacre, 2011).

7.3.1 Rating scale functioning.

It was decided to examine rating scale functioning initially before any other calculations (e.g. Rasch PCA) were performed. This was so that if any items or categories needed to be removed to improve reliability and / or rating scale functioning, such calculations would be performed on an improved measure. This is because if any other calculations are performed first, these would need to be redone after readjusting the rating scale.

After inspection of the 5-point scale functioning, it would appear that categories 2 and 3 needed to be collapsed. This was especially apparent due to the step threshold for 2 (slightly) and 3 (somewhat) not increasing but rather decreasing in category value from -.69 to -.88 (see Table 7.1). Such disordering indicates that this particular category structure (the intended increment from category 2 to 3) is relatively rarely observed by respondents and is therefore unreliable.

Table 7.1. Summary of the RDI Five-Point Rating Scale Category Functioning

Category Label	Observed count	Average measures	In-fit mean square	Out-fit mean square	Step Threshold	Step Standard Error (Structure Standard Error)
1 (not at all)	404	-.96	1.12	1.22	None	None
2 (slightly)	395	-.43	.96	1.01	-.69	.06
3 (somewhat)	832	.12	.97	.94	-.88	.05
4 (very much)	1,187	.78	.94	.84	.11	.04
5 (completely)	1,009	1.97	1.03	1.07	1.46	.05

Note. *Observed count* refers to the total number of responses for a given category.

A probability curve was also produced (Figure 7.1) in which the disordering of the category structure is visibly noticeable. Categories 2, 3 and 4 appear to be the most probable categories and their highest probability of endorsement is about .45. Figure

7.1 shows T1 (at -.69) as the threshold for choosing Category 2 (‘slightly’) over Category 1 (‘not at all) (as also shown in Table 7.1). In the same figure (Figure 7.1) T2 (at -.88) is shown as the threshold for choosing Category 3 (‘somewhat’) over Category 2 (‘slightly’). This indicates a high disordering of category thresholds because the threshold from 2-3 (-.88) is less than the threshold from 1-2 (-.69).

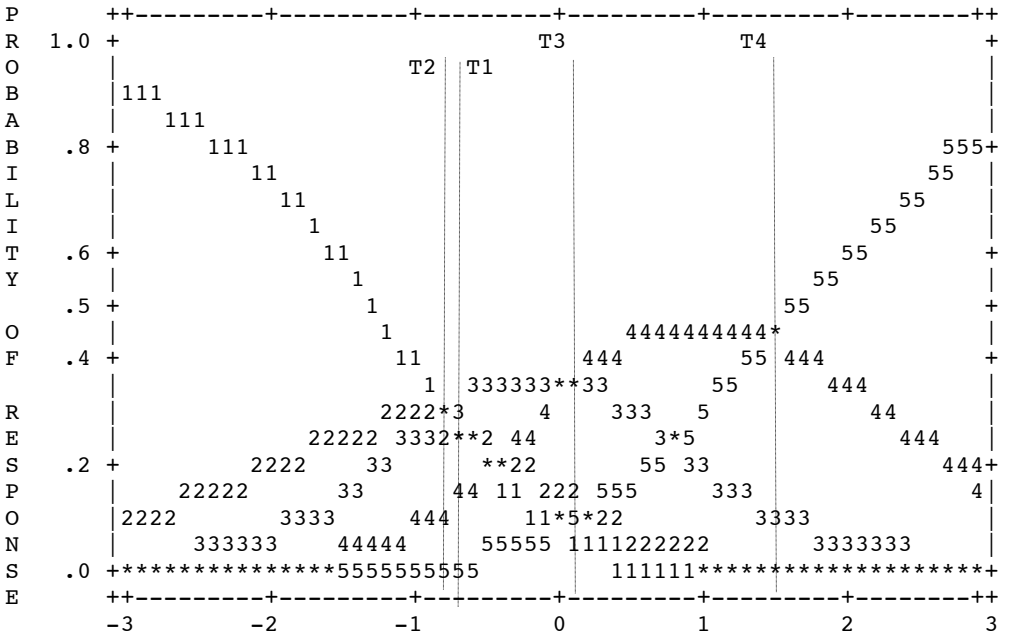

Figure 7.1. Category Probability Curve of the Relational Depth Inventory five-point rating scale.

In Figure 7.1, if a person’s agreeableness was 1 logit lower than the person was able (-1 on the axis), that person’s probability of endorsing a 2 or a 3 is not much different; the probability of endorsing a 2 is about .3 and endorsing a 3 about .25. The probability of endorsing a 4 or a 5 is close to about .1 and .0 respectively.

Before making any final decisions regarding the collapsing of categories in order to address any disorder, various options were explored (see Table 7.2) but the most reliable appeared to be collapsing categories 2 and 3 and removing items 18 (“felt as

if time stopped") and 15 ("felt as if I was going beyond ordinary limits"). Item 15 only became a misfitting item once 18 had been omitted but there were no more misfits once both had been omitted.

Table 7.2. Summary of Changes in Person and Item Separation and Reliability as a Result of Collapsing Rating Scale Categories and Removing Missfitting Items.

	Separation (G)		Reliability				
Rating Scale	Person (G)	Item (G)	Person	Item	In-fit mean square	Out-fit mean square	Number of misfitting items
Original Five-point	3.79	7.93	.93	.98	1.04	1.02	1 (18)
4-point (combining 2 & 3)	3.79	7.75	.93	.98	1.00	.99	1 (18)
4-point (combining 2 & 3, omitting Item 18)	3.78	7.60	.93	.98	1.01	.99	1 (15)
4-point (combining 2 & 3, omitting items 18 & 15)	**3.77**	**7.83**	**.93**	**.98**	**1.01**	**.98**	**0**
3-point (combining 1 & 2, 3 & 4)	3.70	7.09	.93	.98	1.00	1.00	1 (18)
3-point (combining 1 & 2, 3 & 4; Omitting item 18)	3.67	6.95	.93	.98	1.00	.99	0

Note. Emboldened row indicates the best solution, which was used in the study.

In Table 7.2, infit and outfit improved and the average measures and thresholds also improved with category collapsing and removal of items. Clearly, the four-point scale facilitates better person and item separation and so the four-point scale will be adopted statistically for the remainder of the analyses conducted in this current piece of research. This solution improves content validity by removing items (so there are therefore no misfitting items). Four categories also makes the scale simpler to use and there is a minimal cost to separation. See Table 7.3 for summary of 4-point scale and Figure 7.2 for corresponding probability curve (of 4-point scale). Further analyses in the current study, will be calculated by statistically collapsing categories 2 and 3, and statistically removing items 18 and 15.

Table 7.3. Summary of RDI Four-Point Rating Scale.

Category Label	Observed count	Average measure	In-fit mean square	Out-fit mean square	Step threshold	Step standard error
1 (not at all)	332	-1.49	1.08	1.08	None	None
2 (slightly + somewhat)	1104	-.26	.99	1.00	-2.05	.07
3 (very much)	1118	.87	.95	.88	.30	.05
4 (completely)	936	2.40	1.00	1.02	1.75	.05

Note. Omitting items 18 and 15 collapsing categories 2 and 3.

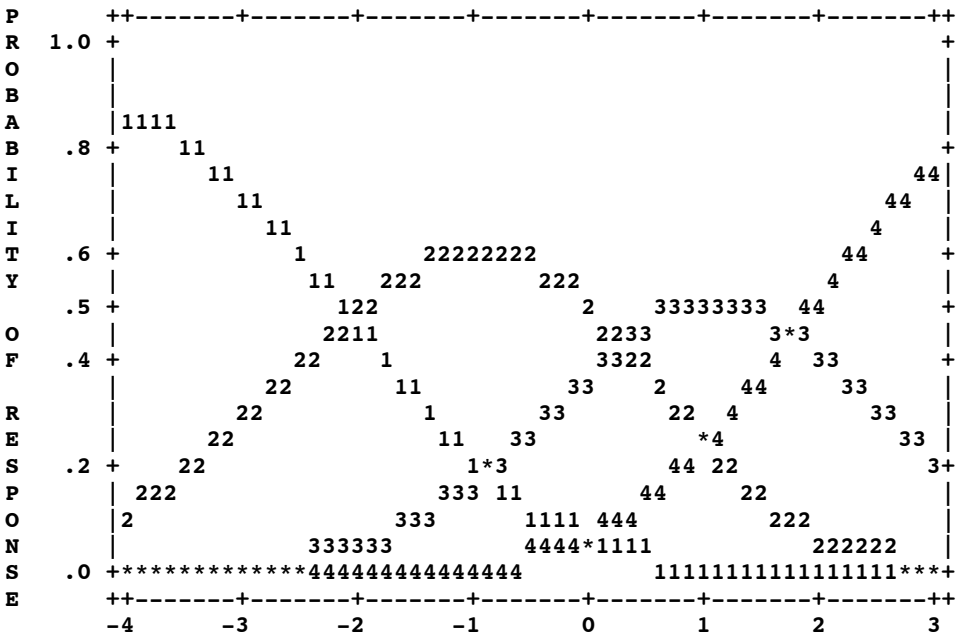

Figure 7.2. Probability curve of the RDI-Revised 4-point rating scale categories (combining 2 and 3, omitting Items 15 and 18). Probability of response categories as a function of perceived relational depth (as assessed by the RDI).

7.3.2 Reliability

Internal consistency. Internal consistency for the 22 items was high (Cronbach's Alpha = .93) indicating that the 22 items as a whole are assessing one construct. There was no indication that removing any more items would improve the measure.

Person reliability index. The Rasch equivalent to internal consistency is *person reliability.* The Rasch person reliability index was also high at .93. This high figure indicates that the persons in this sample are in fact very reliable. In other words, if the same sample of participants were given a different measure assessing the same construct (i.e. relational depth) the same persons would be placed at the same points on the measure as they are in this study.

Item reliability index. The item reliability index was very high at .98 indicating that if the same set of items were given to another sample of persons with the same distribution of strength relational depth presence, items would be placed in the same position along the measure as they are in this study. Put simply, the items are very reliable.

7.3.3 Person and Item Separation and Strata

Item separation index (G). The item separation index was 7.83 indicating that items can be separated into at least 7 distinct groupings of hierarchy. In other words there are at least 7 levels of difficulty in the items.

Person separation index (G). The person separation index was 3.77 indicating there are at least three distinct groups of persons experiencing different strengths of relational depth.

Wright and Masters (1982) state that if statistically distinct levels of item difficulty can be defined as difficulty strata '…with centers three calibration errors apart, then

separation index G can be translated into the number of item strata defined by the test H

$$H = (4G + 1)/3$$

where G = Separation index = "True" standard deviation / Average measurement error…' (p.92). Wright and Masters (1982) also state that the same can be done for persons therefore a strata for persons can be examined. Thus by utilizing the separation indices for items and persons, item strata and person strata can be calculated and mapped.

Item strata H. For the 4-point scale 11 item strata were identified. Where G (item separation index) is 7.83, the item strata is calculated to be 10.77

$$H = (4G + 1) / 3 = (4 \text{ X } 7.83 + 1) / 3 = 10.77.$$

Figure 7.3 shows a keymap showing such item strata. It can be seen that there appear to be approximately 11 different groups or strata of items. The vertical line down the centre of the map is located at a logit value of zero which is indicative of a person experiencing an average strength of relational depth. One can see estimates of what such a person would score on all of the items. For example they would most likely score a 3 on items in the lowest strata and most likely score a 1 or a 2 on items in the top most strata.

As can be seen in Figure 7.3 (which corresponds with Table 7.4), all 11 item strata were named according to which items they were comprised. The four easiest items to endorse form the first stratum and are at the bottom of the diagram (e.g. therapist genuine, there fore me, respect for therapist and I [client] genuine) whereas the most difficult items to endorse were 'spiritual' and 'understanding therapist' which form the last and 11th stratum.

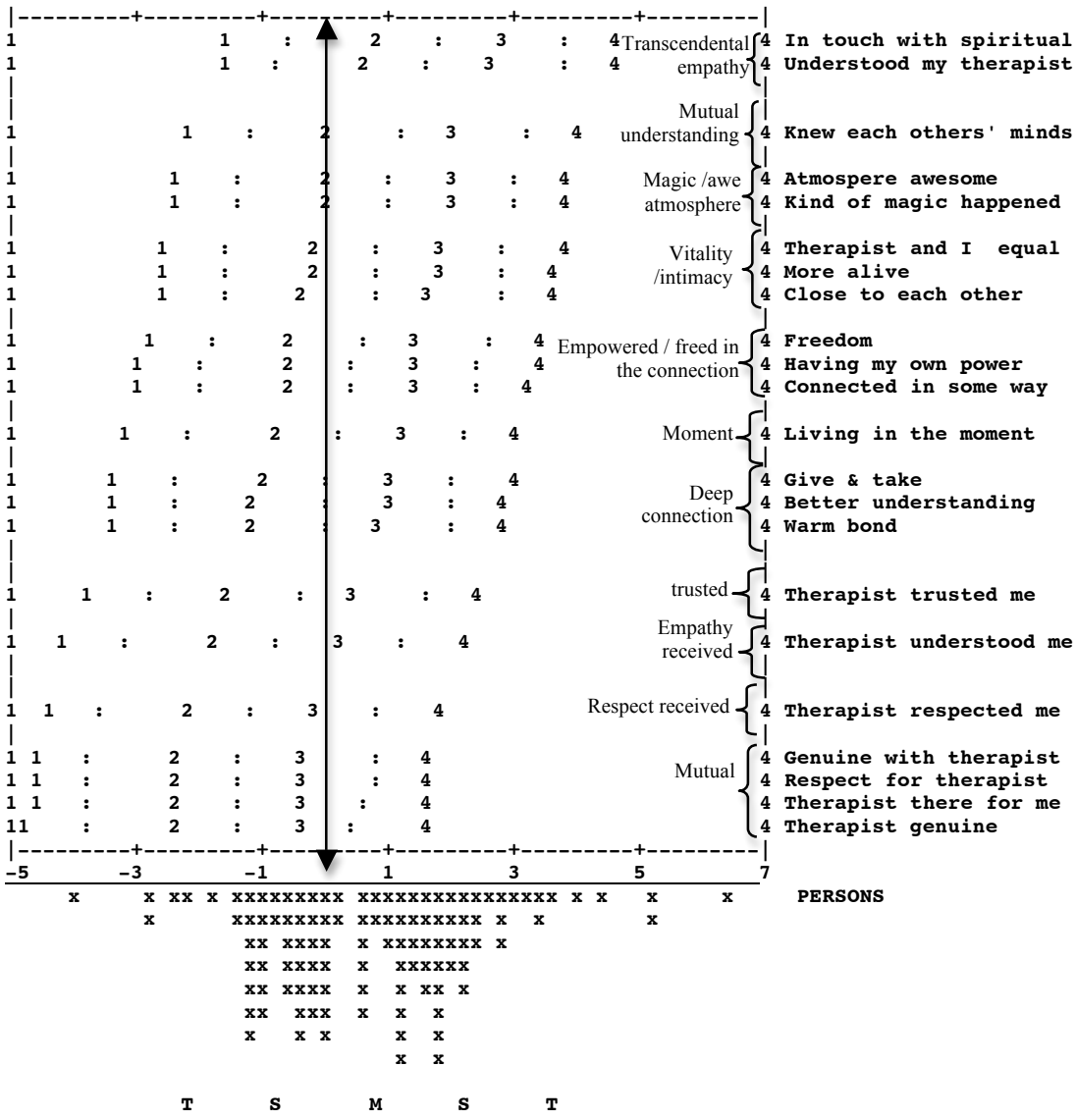

Figure 7.3 Key map showing item strata for 4-category, 22 items measure.

In order to distinguish item strata, measure values for each item were referred to which helped understanding of how statistically the 11 strata are distinguished and how they are different. Table 7.4 (which corresponds with Figure 7.3) shows item measures (showing statistics in measure order showing item boundaries). One can see in the standard error column, that values are .12 or .13. and one can also see the measure column. The difference in the measure values and the standard error both play a part in the discrimination of item strata. For example, the most difficult two items are 'in touch with my spiritual side' with a measure value of 1.68 and 'I understood what it was like for my therapist' with a measure value of 1.56.

After this, the next item (in Table 7.4) (in this case the next strata is only one item) which happens to be 'my therapist and I both knew what was in each other's mind' is more than one standard error different in its measure value which is 1.03. The next two items are quite close in measure values with 'the atmosphere was kind of awesome' having a measure value of .87 and 'a kind of magic happened' having a measure value of .86. These two items, not surprisingly, comprise another strata. These two item complete a strata simply because the next item down is at least one standard error different to this last item. The item 'my therapist and I were equal, which has a measure value of .71, begins the next strata simply because the difference between .86 and .71 is more than one standard error unit difference. As one looks at the table one can see how the strata form in this way.

Table 7.4. Statistics in measure order showing item strata boundaries.

Item	Measure	S.E.	Infit Mnsq	Item-total corr
In touch with my spiritual side (6)	1.68	.13	1.35	.62
I understood what is what like for my therapist (14)	1.56	.12	.93	.66
My therapist and I both knew what was in other's minds (7)	1.03	.12	.95	.60
The atmosphere was kind of awesome (13)	.87	.12	1.08	.70
A kind of magic happened (9)	.86	.12	1.11	.72
My therapist and I were equal (17)	.71	.12	1.40	.58
I felt more alive (8)	.70	.12	.85	.72
My therapist and I felt close to each other (16)	.58	.12	.87	.68
I felt a sense of freedom (1)	.42	.12	1.03	.59
A sense of having my own power (21)	.32	.12	1.06	.62
My therapist and I were both connected in some way (10)	30	.12	.82	.71
I was living in the moment (4)	.08	.12	1.11	.62
There was give and take between my therapist and myself (2)	-.10	.12	.84	.64
I had better understanding of myself and others (23)	-.16	.12	1.32	.57
Warm personal bond as fellow human beings (24)	-.22	.12	.92	.69
Therapist trusted me (11)	-.61	.12	.90	.64
My therapist understood what it was like for me (5)	-.90	.13	1.01	.61
My therapist respected me (3)	-1.27	.13	.81	.63
I was being genuine with my therapist (20)	-1.44	.13	1.03	.54
I felt respect for my therapist (19)	-1.44	.13	.91	.61
My therapist was there for me (22)	-1.45	.13	.89	.62
My therapist was being genuine with me (12)	-1.49	.13	.74	.64

Note. Item strata boundaries are indicated by dotted lines.

Person strata. The person separation of 3.77 indicates that there are about 5 statistically different strata of persons

$$H = (4G + 1) / 3 = (4 \times 3.77 + 1) / 3 = 5.36$$

Figure 7.4 shows a Rasch person-item map and displays such person strata. In this case, a Rasch person-item map can be described as a ruler consisting of the measurements of participants' experience of presence of relational depth and the items assessing such presence. On the left side persons are in an order such that persons experiencing weakest presence of relational depth are at the bottom and those experiencing strongest relational depth are at the top. On the right of the person-item map (Figure 7.4) is the ordering of difficulty of items. At the bottom are the easiest to endorse items and at the top are the hardest to endorse.

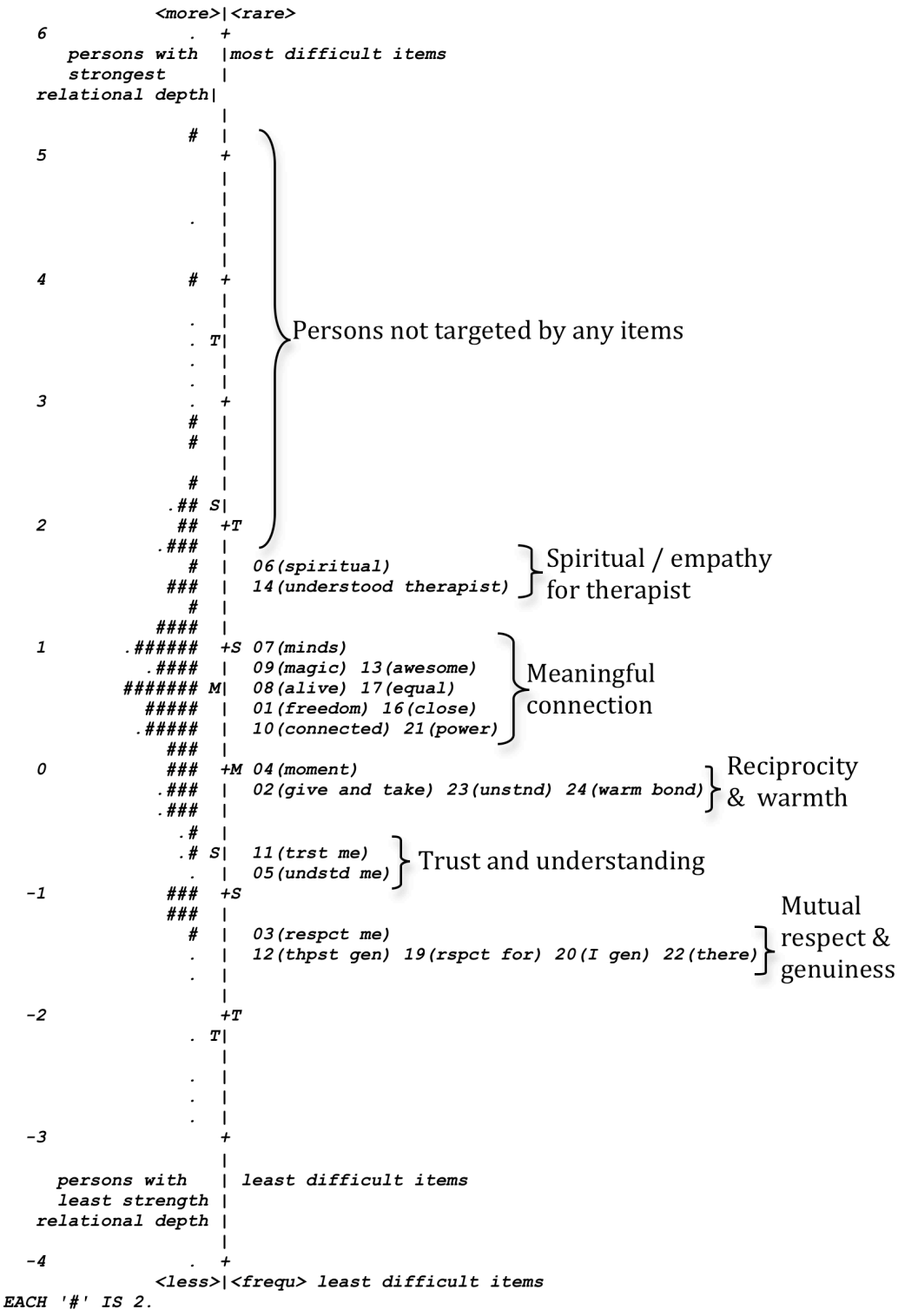

Figure 7.4. Map of persons and items showing 5 different strata of persons. Each # represents 2 persons and every . represents 1 person.

The person measure in the case of this data can be seen as the presence of relational depth of each person and the item measure can be seen as the strength of presence needed to endorse each item. This map of persons and items examines and compares the range and position of the item measure distribution to the range and distribution of the person measure distribution. If presence of relational depth is to be measured by all persons in the sample, items should be as equally spread as persons. More simply, where there are persons on the left side there should be items on the right.

Working one's way up from the bottom of the map (Figure 7.4), from -3.99 logits to -1.72, one can see that 6 persons were not targeted by the items indicating that some items were not easy enough for these persons in the sample. There were no persons from the research clinic sample in this range.

Moving up (again in Figure 7.4), *Mutual respect and genuineness*, the first person stratum situated at -1.48 to -1.30 logits, identified three persons none of which were from the research clinic sample. Such persons were successfully targeted by the least difficult items to endorse which were 'therapist respected me' (Item 3), 'therapist was genuine with me' (Item 12), 'respect for my therapist' (Item 19), I was genuine with my therapist (Item 20) and 'therapist was there for me' (Item 22). These five items are generally concerned with mutual respect and genuineness and this was therefore named as such. The fact that this stratum is positioned lowest indicates that it would seem to identify those persons experiencing very low strength of relational depth. As only three people were in this particular stratum it would indicate that

such items, which assess relatively low levels of relational depth, are targeted by a relatively small number of people.

Between the stratum just previously described and the one described below, from -1.29 to -.92 logits, there appear to be 12 persons not targeted by items. Two of these persons are from the Social Anxiety sample at -1.02 and -.92 logits.

Trust and understanding. Situated at around the -.91 to -.61 logit mark, only four people were targeted by the items in this stratum. Persons in this stratum were targeted by items 'therapist trusted me' (Item 11) and 'therapist understood what it was like for me' (Item 5). This second stratum was therefore named 'trust and understanding.' The positioning of this stratum would indicate that this identifies persons experiencing lower to mid strength of relational depth and that such levels of relational depth include those persons who experience their therapist trusting them and understanding them. No research clinic participants were targeted by this strata.

From logits -.60 to -.22 ten persons were not targeted by any items. Again, this is of a fairly narrow range and is therefore of little concern. Two persons were from the research clinic sample, one of them from the social anxiety sample.

The third stratum*, Reciprocity and Warmth*, targeted twelve persons from logits -.21 to .10. The 13 items in the strata were 'I was living in the moment' (Item 4),'give and take' (Item 2), 'better understanding of self and other' (Item 23) and 'warm personal bond as fellow human beings' (Item 24). This stratum identifies persons

experiencing reciprocity and warmth in particular and indicates that persons experiencing these qualities are experiencing moderate strength of relational depth. Five of the persons were from the research clinic sample, one of which was from the social anxiety sample.

Before the fourth stratum there were 6 persons not targeted by items. Such person ranged from .11 logits to .18 logits which is of little concern because it is such a small range.

Meaningful Connection, the fourth item stratum contained the largest number of persons of all the five strata. Ranging over the distance of nearly one logit, from .19 logits to 1.13 logits, this stratum identified 57 persons within the space of 1 logit distance. This stratum's range included the mean for persons indicating that the average person was in this stratum. A total of nine items targeted such persons with items assessing 'having own power' (Item 21), 'feeling connected' (Item 10), 'sense of freedom' (Item 1), 'feeling close to each other' (Item 16), 'feeling more alive' (Item 8), 'feeling equal with therapist' (Item 17), 'magic' (Item 9), 'awesome atmosphere' (Item 13) and 'a sense of knowing each other's minds' (Item 7). The position of this particular stratum would indicate persons identifies by this stratum experience moderate to high strength relational depth which is characterized by a meaningful connection with the therapist and includes experiences of feeling empowered, connected, sense of freedom, close to therapist and magical feelings. A total 16 persons were from the research clinic sample with 7 coming from the social anxiety portion of the sample.

Persons situated from 1.14 to 1.43 logits on the measure were not targeted by any items. A total of 10 persons were not targeted here.

Spiritual / Empathy for therapist. The fifth and final stratum ranged from 1.44 to 1.75 logits targeted 8 persons with just two items; 'in touch with spiritual side' (Item 6) and 'felt I understood what it was like for my therapist' (Item 14). This last stratum's high position on the map indicates that it targets those persons who experience high levels of relational depth and that experiences of feeling spiritual and having empathy for the therapist typically accompany such high levels of relational depth. These items were therefore the most difficult to endorse.

Towards the top of the person-item map one can see that on the left side there is a considerable spread of persons which are not targeted by any items. On the measure this is from about 2 to 5 logits. This would indicate that there is some level of inadequacy in the measure in that exceptionally high levels of relational depth are not assessed. Put simply, none of the questionnaire items are difficult enough for persons across this range of the measure.

To summarise, 85 persons were targeted by items with 77 persons not targeted. There appeared to be no pattern regarding any of the research clinic sample, including the social anxiety portion.

7.3.4 Fit Statistics

Table 7.4 gives item logits in difficulty order and includes standard errors infit statistics and item total correlations. Since none of the items' infit values are above 1.44 and none of the item total-correlations are below .4, evidence suggests that all items are functioning similarly and are consistent with each other.

7.3.5 Rasch PCA

The total variance explained by the measure was 53.9% and the unexplained variance was therefore 46.1%, equating to 22 units (eigenvalues). There was evidence of two contrasting dimensions. The eigenvalue of the first contrasting dimension is 3.1 (rounded to 3), which indicates it has the strength of about 3 items. If a contrasting dimension is two units or more it can be considered to be a dimension (Linacre, 2011). As this is bigger than the strength of two items, this indicates that there is a contrasting dimension. This largest contrasting dimension explains 6.6% of the total variance. This is larger than the 4% that would normally be observed for data to fit the Rasch model. See Table 7.5 for Rasch PCA residual results.

Table 7.5 Rasch PCA residual (unexplained variance statistics).

	Units/Eigenvalue	Percent of total variance	Percent of unexplained variance
Total explained variance	25.7	53.9%	-
Total unexplained variance	22.0	46.1%	100%
Unexplained variance in 1st contrast	3.1	6.6%	14.3%
Unexplained variance in 2nd contrast	2.8	5.8%	12.7%
Unexplained variance in 3rd contrast	1.7	3.6%	7.8%
Unexplained variance in 4th contrast	1.5	3.2%	7.0%
Unexplained variance in 5th contrast	1.4	3.0%	6.6%

The eigenvalue of the second contrasting dimension is 2.8 (rounded to 3) which, like the first, is more than two (also indicating it to be considered a dimension); it has the

strength of about 3 items and explains 5.8% of the total variance. These two contrasting dimensions, account for 12.4% of the total variance. A factor sensitivity ratio was then calculated by summing the eigenvalues of this first and second contrast and dividing the result by the explained variance units. The first contrasting factor accounted for 3.1 units (eigenvalues) and the second 2.8, totaling 5.9 out of a total of 25.7 explained variance units (eigenvalues). Therefore 5.9 / 25.7 = 22%. This indicates that 22% of the measure's stability is affected by two contrasting dimensions in the responses. Interpretation of both contrasting dimensions will be done by examining Rasch PCA contrast plots. Linacre (2011) states that in the interpretation of Rasch PCA, the contrast between the positive and negative loadings is key to interpreting Rasch PCA. Therefore interpretation of contrasting dimensions will also include factor loadings tables that display such loadings, using a loading ≥ .4 as an interpretive criterion. Rasch PCA plots and loadings tables give information as to how items differ from the measure, in other words, a measure of their contrast. However, Rasch PCA plots also plot items with regard to their difficulty; an X axis plots the measure (item difficulty) and a Y axis plots the contrast, which is essentially the distance from the measure (which correspond with the factor loadings table). These Rasch PCA plots are shown in Figures 7.5 and 7.6 for the first and second contrasting dimensions respectively. Such loadings are displayed in Tables 7.6 and 7.7 for the first and second contrasting dimensions respectively.

Interpretation of First Contrasting Dimension. Combining information from the factor loadings table (Table 7.6) and the Rasch PCA plot (Figure 7.5), the furthest positively loaded items appear to have a common meaning of *focus on self with self*

and the negatively loaded items appear to have a common meaning of *focus on self with therapist.* Therefore the first contrasting dimension evidences that there may well be some relatively minor influence that might be aptly named "focus on self."

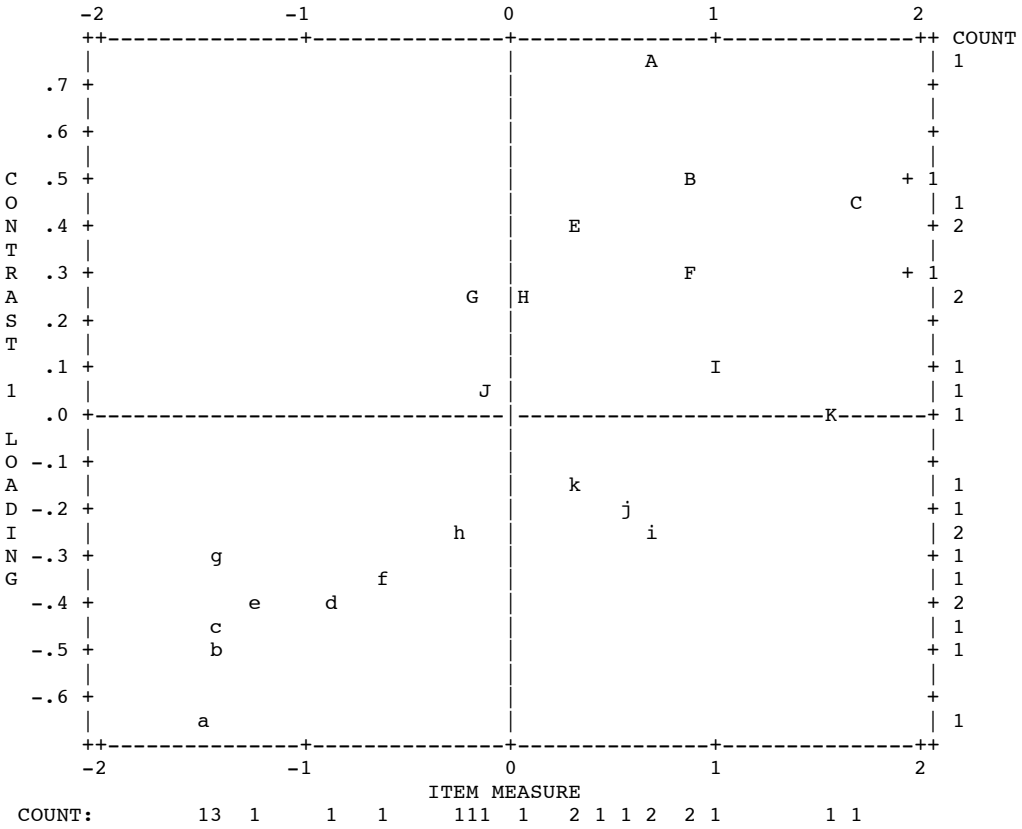

Figure 7.5. Rasch PCA plot of first contrasting dimension. The display plots the measure on the X axis against the contrasting dimension on the Y axis.

Second Contrasting Dimension. As with the first contrast, a Rasch PCA plot (Figure 7.6) and a factor loadings table (Table 7.7) will be used to interpret findings. The positively loaded items given on the loadings table appear to have a common meaning of *enmeshment* and the negatively loaded items appear to have a common meaning of *differentation.* Therefore, this second contrasting dimension could be named as "interdependence."

Table 7.6. Factor loadings for principal components of first contrasting dimension.

Item	Factor Loading
I felt more alive (8) =A	**.73**
I felt a kind of magic happened (9) =B	**.51**
I felt in touch with my spiritual side (6) =C	**.44**
I felt a sense of freedom (1) = D	**.40**
I felt a sense of having my own power (21) = E	.39
I felt the atmosphere was kind of awesome (13) = F	.29
I felt I had a better understanding of myself and others (23) = G	.27
I felt I was living in the moment (4) = H	.24
My therapist and I both knew what was in each other's mind (7) = I	.10
There was give and take between my therapist and myself (2) = J	.06
I felt my therapist was being genuine with me (12) = a	**-.64**
I felt respect for my therapist (19) = b	**-.51**
I felt my therapist was there for me (22) = c	**-.46**
I felt my therapist understood what it was like for me (5) = d	**-.40**
I felt my therapist respected me (3) = e	-.38
I felt my therapist trusted me (11) = f	-.36
I felt I was being genuine with my therapist (20) = g	-.31
I felt a warm personal bond between myself and my therapist as fellow human beings (24) = h	-.26
I felt my therapist and I were equal (17) = i	-.26
My therapist and I felt close to each other (16) = j	-.20
I felt my therapist and I were both connected in some way (10) =k	-.15
I felt I understood what it was like for my therapist (14) = K	-.02

Note. Item in bold denote items used in contrasting dimension

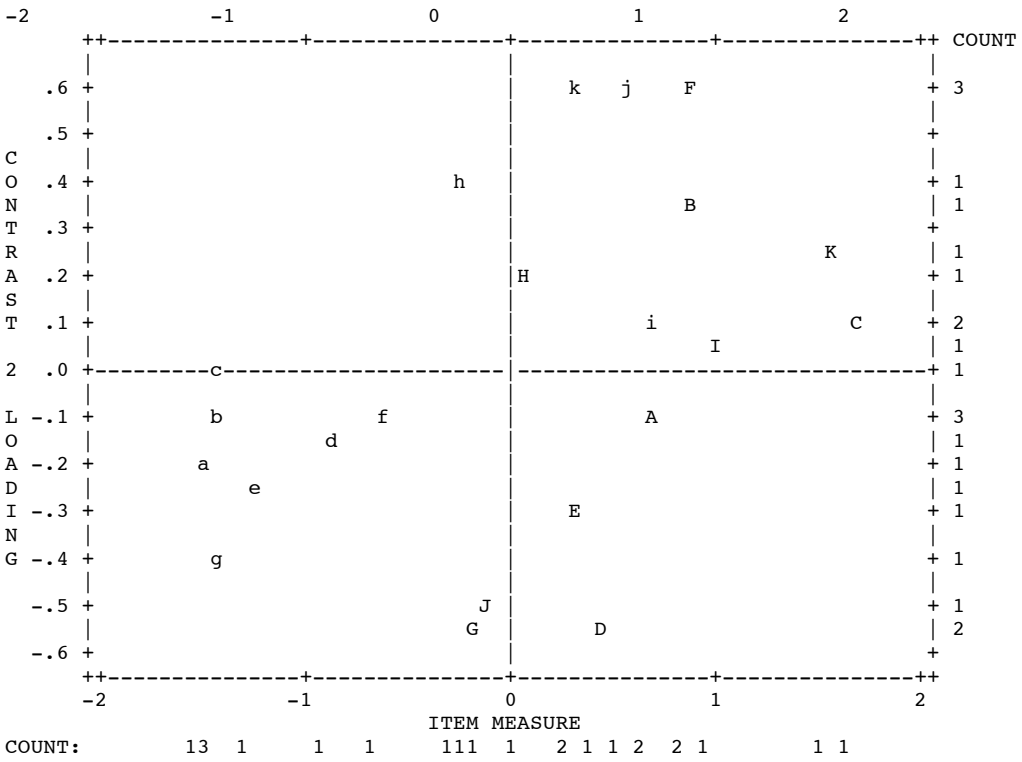

Figure 7.6. Rasch PCA plot of second contrasting dimension. The display plots the measure on the X axis against the contrasting dimension on the Y axis. Please see Table 7.7 for coding of items

These two contrasting dimensions each with sets of residual variables thus provide some evidence for two secondary scales within the RDI-R. These two contrasting dimensions are small compared to the overall relational depth variable, which accounts for more than four times as much variance as these two put together; the first contrast accounts for 6.6% of the total variance and the second accounts for 5.8% totalling 12.4% of the total variance. These results would suggest then that the RDI-R is slightly multidimensional but that this multidimensionality is relatively minor.

Table. 7.7. Factor loadings for principal components of second contrasting dimension.

Item	Factor Loading
My therapist and I felt close to each other (16) = j	**.60**
I felt the atmosphere was kind of awesome (13) = F	**.59**
I felt my therapist and I were both connected in some way (10) =k	**.59**
I felt a warm personal bond between myself and my therapist as fellow human beings (24) = h	**.40**
I felt a kind of magic happened (9) =B	.36
I felt I understood what it was like for my therapist (14) = K	.24
I felt I was living in the moment (4) = H	.21
I felt my therapist and I were equal (17) = i	.11
I felt in touch with my spiritual side (6) =C	.08
My therapist and I both knew what was in each other's mind (7) = I	.05
I felt a sense of freedom (1) = D	**-.57**
I felt I had a better understanding of myself and others (23) = G	**-.56**
There was give and take between my therapist and myself (2) = J	**-.50**
I felt I was being genuine with my therapist (20) = g	**-.40**
I felt a sense of having my own power (21) = E	-.31
I felt my therapist respected me (3) = e	-.26
I felt my therapist was being genuine with me (12) = a	-.19
I felt my therapist understood what it was like for me (5) = d	-.17
I felt my therapist trusted me (11) = f	-.12
I felt more alive (8) =A	-.10
I felt respect for my therapist (19) = b	-.09
I felt my therapist was there for me (22) = c	00

To further explore the extent to which the eight significantly-loading items were off-dimension, residual correlations were explored. Table 7.8 shows such correlations. On inspection of Table 7.8, the on-dimensional item (13, atmosphere awesome) and the off-dimensional item (9, kind of magic happened) correlate at .42. An

interpretation of this is that the larger amount of residual variance that exists in the off-dimensional item is correlating with the slightly smaller amount of residual variance in the on-dimensional item. This indicates that the on-dimensional item is actually close to being off-dimension. When this is checked out with the factor loadings of residuals this is confirmed because the on-dimensional item is close to being off; it's loading is .29 and therefore it is not far from being off-dimension. The Rasch PCA plot also helps which shows both these items are located at the same point on the measure but at a different point on the contrasts but not a huge difference on the contrasts (a difference of 2 logits which is not huge given that the biggest contrasting items have a difference of 7 logits).

Table 7.8. Largest observation residual correlations used to identify dependent items

Corr	Item	Item
.42	**9. I felt a kind of magic happened**	13. Atmosphere awesome
.37	11. I felt my therapist trusted me	**12. Therapist genuine**
.35	21. I felt a sense of having my own power	23. Better understanding
.35	**8. I felt more alive**	**9. I felt a kind of magic**
.34	10. Therapist and I connected	16. Therapist and I close
.33	**1. Sense of freedom**	**8. Felt more alive**
-.40	**1. Sense of freedom**	10. Therapist and I connected
-.37	**8. I felt more alive**	**12. Therapist was being genuine**
-.35	16. Therapist and I close to each other	23. Better understanding
-.32	**1. Sense of freedom**	16. Therapist and I close.

So this is why these items' residuals correlate; their pattern of residuals happen to be similar and they are at the same place on the measure (making their residuals correlate); their contrast is different enough to make one on-dimension and one off-dimension but not so different that their residuals have no association at all.

A slightly different interpretation can be made for the negatively correlated items 'sense of freedom' (Item 1) which is off-dimension and, 'therapist and I connected in some way' (Item 10) which is on-dimension. They are at the same point along the measure at about .3 logits but differ in their contrast from the measure. Item 1 is +4 logits away from the measure and Item 10 is about -2 logits away. This is a difference of about 6 logits apart from each other. The difference in signs between these two items explains the negative correlation; in other words each item contrasts in a different direction to the other. Just like the correlation mentioned previously, this indicates their pattern of residuals happen to be similar and they are at the same place on the measure (making their residuals correlate) but their contrast is different enough to make one on-dimension and one off-dimension but not so different that their residuals have no association at all. However, in spite of there being some moderate correlation of residuals, none appear to be highly locally dependent in that none of the correlations were above .7. Therefore all items are considered to be locally independent enough that none are considered to violate this key assumption of the Rasch model.

7.3.6 RDI and WAI targeting of persons

In order to explore how the RDI-R and the WAI targeted persons, a person-item map was produced that included both WAI and RDI items. This is shown in Figure 7.7. Unlike previous analyses, only 43 participants (61 cases) were included as the WAI was not completed by web-based participants. On inspection of the person-item map, it would appear that the RDI targets a larger spread of people that the WAI. Furthermore, the RDI targets persons higher on the measure (the measure here being

a combination of WAI and RDI). This would indicate that the RDI in fact provides an upward extension of the client-therapist relationship not assessed by the WAI.

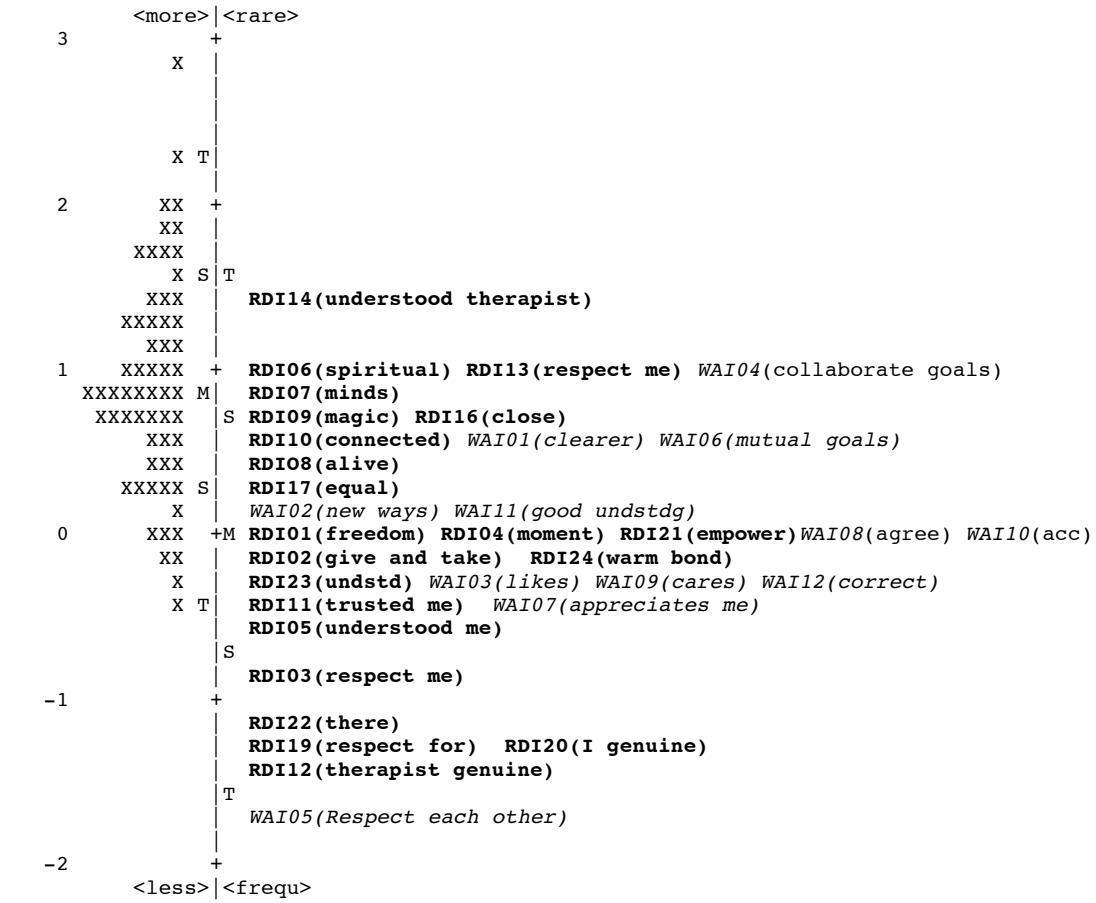

Figure 7.7 Person Item map of RDI and WAI Items. WAI items are in italics

7.4 Chapter Summary

In Study 2, extensive analyses based on the Rasch Model provided considerable construct-related evidence in that the RDI-R is largely unidimensional. Study 2 was also informative in that it indicated what experiences might characterise different strengths of relational depth. It also suggested that the RDI-R did not target persons experiencing very strong levels of relational depth. The measure as a whole evidences excellent reliability and did so even before the rating scales were adjusted.

Added to this, internal consistency was also excellent, the measure appears to be sound in this respect.

The item hierarchy indicated that there are 11 distinct groups (strata) of items on the RDI, with each strata containing items of roughly equal difficulty. Items easiest to endorse assessed those experiences that concerned the client's experience of the therapist. The moderately difficult to endorse items assessed aspects of the relationship, while the harder to endorse items included items that assessed the client understanding the therapist as well as spiritual experiences. At the time of writing, there is no specific theory on relational depth that looks at which relational depth experiences may be stronger, weaker or more difficult to experience. However, there is previous research which suggests that there may be different strengths or different aspects of relational depth. In a qualitative study, Knox (2008) found that out of three aspects of clients' experiences of a moment of relational depth (self, therapist and relationship) it was the relationship aspect where clients reported to have 'deeper' feelings. Clients also reported that they felt the experience of the relationship was on 'a different dimension' (compared to experience of self and therapist) and that experiences of feeling spiritual, mystical and mutual understanding accompanied this. Such findings are consistent with the item hierarchy in the current study, where items assessing spiritual experiences and the client understanding the therapist were most difficult to endorse.

The Rasch principal components analysis also provided evidence that all the items fit onto the dimension, but some appear to some extent to be partially measuring

something else. However, the results of the PCA indicated that the measure was largely unidimensional and that the 22 remaining items assessed the main dimension of relational depth.

Nevertheless, the evidence suggests that there were two, much smaller, contrasting dimensions; "focus on self" and "interdependence." The fact that subordinate dimensions emerged like this, within one data set, gives credence to, In the current findings, items assessing feeling more alive, 'a kind of magic', spiritual, and a sense of freedom are clearly experiences involving the self focusing on self and items assessing the therapist being genuine with self, respect for therapist, therapist there for self and therapist understanding what it is like for self are clearly experiences of the self focusing on the self in relation to the other. Furthermore, for the second contrasting dimension in this study (interdependence: enmeshment vs differentation) items here assessed therapist and self as feeling close and being connected (enmeshment) and also assessed (sense of freedom, better understanding of self and others and give and take (differentation). Clearly, therefore, the two contrasting dimension found in the current study are consistent with the Structural Analysis of Social Behaviour (SASB) model (Benjamin, 1996, 2005).

How many distinct groups of clients are there who experience relational depth?

Five distinct groups were evidenced. Item difficulty determined the groupings of persons. For example, the first grouping was named 'mutual respect and genuineness' and items assessing such experiences were easiest to endorse. This was followed, in the order of difficulty by 'trust and understanding', 'reciprocity and

warmth', 'meaningful connection' and 'spiritual / empathy for therapist.' This supports the idea that clients can be grouped by different strengths of relational depth.

How well does the RDI-R target persons; are there persons it does not target?

However, there was evidence that a substantial number of clients over a wide range of different strengths of relational depth were not targeted sufficiently by items on the measure. This non-targetting was most prominent in the higher scoring persons suggesting that their experience of relational depth was stronger than there were items for. This could imply that there is a quality of relational depth that is perhaps beyond words. Therefore in any future measure, there may need to be a selection of items designed to assess higher strengths of experience. This may be difficult as such experiences may be beyond language and it may not be possible to create items that assess very strong experiences of relational depth.

Does the WAI target persons in a similar or different way to the RDI?

It would appear that the RDI assesses working alliance and something else that appears to be an upward extension of the relationship between client and therapist. This would imply that the RDI is not only a measure of relational depth but also a measure of working alliance. It could also imply working alliance is a prerequisite of relational depth.

The results are further discussed in more detail in Chapter 9, the main discussion section.

Chapter 8: Study 3: Process-Outcome Study

8.1 Introduction to Chapter

This chapter documents Study 3 which continues on from the Study 2. As the RDI-R had been tested for reliability and validity in Study 2 the next logical step forward was to use the RDI-R in a process-outcome study. As the RDI-R showed considerable evidence for validity and reliability, the following study sought to use it in order to explore the relation between relational depth and outcome. The following study, therefore, carries on from the previous study in that the RDI-R is used to assess whether presence of relational depth during a significant event is related to positive therapeutic outcome.

The current study had one main aim which was to explore two models of the contribution of working alliance and relational depth to therapeutic outcome over and above pre-therapy client functioning. In the first model, I wanted to explore the unique contribution of working alliance (as assessed by the WAI) whilst controlling for pre-therapy and relational depth as assessed by the RDI. In the second, I wanted to explore the unique contribution of relational depth (as assessed by the RDI-R), over and above pre-therapy scores and the WAI together. This would have the effect of testing whether the RDI contributed significantly to outcome after controlling for pre-therapy functioning and working alliance. Figure 8.1 visualizes what the two different orders mean conceptually. In Model 1, Pre-therapy score and Relational Depth are controlled for in order that the single contribution of Working Alliance may be calculated. Similarly in Model 2, Pre-therapy and Working Alliance are

controlled for in order that the single contribution of Relational Depth may be calculated.

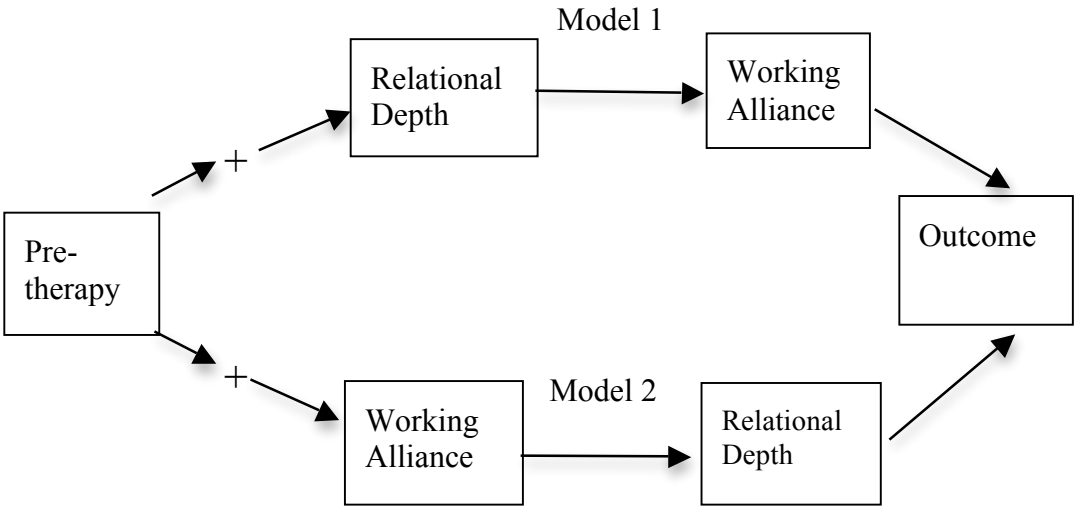

Figure 8.1 Two models of prediction of outcome.

8.2 Method

8.2.1 Participants

Forty-two clients were recruited from the Strathclyde Research Clinic (this had orginally been 43 but one set of data was unusable due to lack of outcome data). Regarding gender there 15 males (36%), 24 females (57%) and 3 (7%) where gender was unknown.

Fifteen (36%) of the clients were suffering from social anxiety (SA) as assessed by the Structured Clinical Interview for DSM-IV Disorders (SCID-IV) and 27 (64%) were not diagnosed with any disorders. The clients not diagnosed with any disorders shall be referred to as Practice-Based (PB) clients simply to differentiate them from

the SA clients (even though all clients were effectively practice-based). The average number of sessions for each client was 28 with the minimum number being 3 and the maximum 67. Of the 42 clients, 23 had ended therapy and 19 clients were at mid-therapy but had attended at least 10 sessions. The therapy approach used was either person-centred or emotion-focused; clients received one or the other but exact numbers are not known due to missing data.

8.2.2 Measures

Relational Depth Inventory – Revised (RDI-R). The RDI-R was used in the current study. The RDI-R consists of a question that asks participants to describe a helpful event in therapy and then asks them to rate that event using 24 items. Each of the 24 items was presented with a Likert scale with scores ranging from 1 to 5 (1 = not at all, 2 = slightly, 3 = somewhat, 4 = very much, 5 = completely). In light of the previous Rasch analysis study, after data was collected, Likert scales were statistically collapsed to a 4-point scale and two items removed. (Please see results section under sub-heading 8.3.1 Data for more information). A 4-point scale could not be presented at the time of data collection because the raw data for the current outcome study was part of the raw data for the Rasch study – the data used for both studies was collected simultaneously. After data was collected, Examples of items include 'I felt my therapist respected me' and 'I felt my therapist and I were both connected in some way.' Please see Appendix 3 for the complete RDI-R.

Working Alliance Short Form-Revised (WAI-SR). The revised short version of the Working Alliance Inventory (WAI-SR; Hatcher & Gillaspy, 2006) was added in order to explore the relationship between working alliance and those events

characterised by relational depth. This version of the Working Alliance was the latest short version and consisted of 12 items. Each item utilised a 5-point Likert scale instead of the usual 7-point scale used in previous short versions. There were 4 items addressing Bond, 4 addressing Tasks and 4 addressing Goals. The WAI-SR items were presented after the 24 relational depth items to clients only.

Structured Clinical Interview for DSM-IV Disorders (SCID-IV). For clients to be deemed as having social anxiety, they had to meet the diagnostic criteria for Social Anxiety, according to the Diagnostic and Statistical Manual Fourth Edition (DSM-IV). The Structured Clinical Interview for DSM-IV (SCID-IV) is an operationalization of the DSM-IV and such interviews were carried out at the research clinic in order to assess clients' social anxiety.

CORE. The Clinical Outcome Routine Evaluation–Outcome Measure (CORE-OM) (Barkham et al, 1998) was used in the current study as one of the therapeutic outcome measures. The CORE-OM is designed to assess the psychological distress of clients during pre-therapy, mid therapy and end of therapy. The CORE-OM consists of 34 items that address three main areas of psychological distress in clients: 4 items assess absence of subjective well-being (e.g. 'I have felt ok about myself'), 12 items assess problems/symptoms (e.g. 'I have felt unhappy') and 12 items assess life functioning (e.g. 'I have thought I have no friends'). A fourth set of 6 items also assess risk/harm (e.g. I made plans to end my life).

Designed to increase sensitivity, half the items focus on low intensity problems (e.g.

'I feel anxious/nervous') and half focus on high intensity problems (e.g. 'I feel panic/terror'). Further, 25% of the items are positively framed. Items are presented and scored on a 5-point rating scale and scored from 0 to 4 (0=*Not at all*, 1=*Only occasionally*, 2=*Sometimes*, 3=*Often*, and 4=*All or most of the time*). In order that clinically meaningful differences can be detected, CORE clinical scores are calculated by calculating the mean of all items, and then multiplying this mean by 10. This means that a total clinical CORE score can range from 0 to 40. For completed questionnaires to be considered valid there must not be more than 3 items omitted (Evans et al, 2002).

There are a number of reasons for using the CORE-OM in the present study. Firstly, it is a well-known and widespread method of assessing outcome. Because it is such a widely used measure there is considerable data which exists about the measure's properties. For example, internal consistency reliability for the 34-item scale in clinical (n = 713) and non-clinical (n = 1009) samples have been found to be .94 and .94, respectively (Barkham et al., 2001). Furthermore, it has been found to correlate highly with the Becks Depression Inventory (alpha =.85) and had good test-retest reliabilities (Evans et al., 2002).

For the same reasons, a number of indices (or cut-off points) have been made available. For example, a mean score of more than 1 (equivalent to a clinical score of more than 10) indicates the patient or client is in the clinical population. This enables the researcher or practitioner to calculate whether clients or patients move from the clinical to the non-clinical population after therapy has completed.

Another benchmark or cut-off point includes the calculation for reliable change from pre to post-therapy. If the mean score from pre to post therapy decreases by .5 or more, it is considered to be a reliable change.

However, there is evidence that the factor structure of the CORE-OM does not reflect the four subscales as intended in that most have high correlations with each other; that three of the subscales (not the risk subscale) would be most constructively used if they were combined (Lyne, Barrett, Evans & Barkham, 2006). In addition, Murphy (2010) states that the risk items are also able to be used to indicate client distress. Therefore, in the current study, the CORE-OM subscales will not be used but rather mean scores will be calculated for each participant and a reliable change index as well as pre and post clinical index scores will be calculated for each client.

Strathclyde Inventory

The Strathclyde Inventory (SI) (Freire & Cooper, 2007) is an outcome measure designed to assess the psychological well-being of the client. It is a self-report instrument that consists of 31 items developed according to Rogers' (1961) description of the 'fully functioning person' (p.183). Based on previous research with a non-clinical population, the internal reliability is high, and factor analyses identify two factors (congruence/fluidity and incongruence/structure-boundness), and convergence with CORE-OM and IIP-26 (Inventory of Interpersonal Problems) is moderately high. The SI has been found to be sensitive to change in psychotherapy and shows substantial pre-post effects (Freire and Cooper, 2007).

Personal Questionnaire (PQ)

The Personal Questionnaire (PQ) (e.g. Elliott, Shapiro and Mack, 1999: Wagner and Elliott, 2001) is a different kind of outcome measure from those mentioned already. It is different in that it is the client who creates their own items which are essentially issues that the client wishes to work on or address during therapy. This is usually constructed with the help of an interviewer who facilitates the client's decision-making rather than directing it, although suggestions are made in order to help the client. In generating items it is suggested to the client that the items they create should be the most important to them. It is also suggested that the client may want to look at their symptoms, specific performance/activity (e.g. work), relationships, self-esteem and emotions and inner experiences. The interviewer helps the client clarify their items and may suggest that goals are rephrased into problems. This is to prevent the client from forming too many unrealistic goals and helps the client focus on problems rather than focusing on what might not be achievable. It is better that items present problems or difficulties that they are experiencing. Such items may begin with "My thinking", "I feel", "I am", 'I can't", and so on. According to Elliott (Elliott, Shapiro & Mack, 1999), a good PQ item would include characteristics that reflect an area of difficulty, is something that the client wants to work on in therapy, refers to a specific, single problem, is in the client's own words, and is not redundant with another PQ item (is not too similar). Each item (problem) is on a 7-point scale and the client rates how much the problem has bothered them during the last seven days (a score of 1=not at all, 2=very little, 3=little, 4=moderately, 5=considerably, 6=very considerably, 7=maximum possible). For data analysis the mean score is calculated and then compared at meaningful points in therapy (here, pre and post therapy).

8.2.3 Procedure

Advertisements were displayed in local GP surgeries as well as local supermarkets asking for the participation of those who wanted or needed therapy. After screening interviews, clients commenced therapy. Each client completed one of each of the outcome measures (CORE, SI and PQ) which were part of an outcome pack. Outcome packs were given to clients prior to change interviews, which took place before therapy (referred to as pre-therapy stage), at every 10th session and at the end of therapy (referred to as post-therapy). Change interviews were conducted, in confidence, by an interviewer. The outcome packs were given or sent out to clients at every 10 sessions (e.g. at 10th, 20th and 30th session) in order that they would be completed in time for each clients change interview which typically took place a few days after the client's 10th, 20th, 30th etc therapy session. Therefore outcome packs were usually completed away from the research clinic and would have been completed wherever the client wanted to complete them. RDI-Rs were administered as part of such outcome (packs with the exception of the pre-therapy stage). Working Alliance Inventories (WAIs), on the other hand, were not completed as part of the outcome packs but completed at the research clinic immediately after sessions 3, 5, 10 and then every 5th session after that. This means that, for every client, there would be one completed pre-therapy outcome pack, one completed post therapy outcome pack (with an RDI-R) after every 10 sessions and one completed WAI after every 10 sessions.

8.3 Results

8.3.1 Data

The previous (Rasch analysis) study found the RDI-R to be more reliable, with two items omitted (15 and 18) and with two Likert points collapsed. Consequently, such items were statistically removed and Likert scales statistically collapsed before any further analyses began. Therefore, although the RDI-R has 24 items and a five-point Likert scale, statistically a 22-item 4-point Likert scale was used in the analyses presented here. This was provide greater integrity and reliability in such analyses. Once such omissions and collapsing had been conducted, participants' mean scores were calculated for pre-therapy and post-therapy on all three outcome measures. RDI-R post-therapy scores were totalled and mean scores were also calculated. Lastly, WAI scores at last session (or nearest to last session) were totalled and mean score calculated.

8.3.2 Data Screening & Tests of Normality

Before any correlation or regression analyses were conducted, all variables were screened for normality assumptions. Skewness is the degree to which a data set is asymmetrical and kurtosis is the degree to which a data set's distribution is flat or steep when compared to a normal distribution (Howitt & Cramer, 2011). Therefore, skewness and kurtosis statistics are indicative of the normality of the spread of the data (Tabachnick and Fidell, 2007). If the data are perfectly normally distributed, the z-scores of skewness and kurtosis should be zero. However, in reality, virtually no data sets are perfect and a z-score not exceeding 2.58, for a data set of this size, would be considered to be an insignificant amount of skewness and kurtosis. The figure of 2.58 is consistent with the significance level of $p < .01$. The variables' z-

scores ranged from .10 – 1.47 and therefore no variable had any significant skewness or kurtosis. In order to test for multivariate outliers, Mahalanobis distance scores were calculated for all regression analyses reported here. No outliers were detected in that none of the distance scores exceeded the critical value score.

8.3.3 Clinical and Reliable Change

Reliable and clinical change was calculated for all three outcome measures. In the case of CORE and the PQ, improvement is indicated by a reduction in scores; a mean score reduction of .5 for CORE scores and a 1.43 reduction for PQ score is indicative of reliable change and a significance level of $p > .05$. For the SI an increase in score is indicative of improvement and an increase in mean score of at least .6 indicates reliable improvement. For CORE, clinical populations are considered to be those with a mean score of more than 1, for the SI, a score of less than 2.45 and for the PQ, a score of more than 3.5. Table 8.1 shows mean outcome scores for pre and post therapy with effect sizes and numbers of participants with clinical and reliable change. Post-therapy refers to the latest point at which participants completed outcome packs. Of the 42 clients, 23 had ended therapy and 19 clients were at mid-therapy but had attended at least 10 sessions.

8.3.4 Preliminary Correlation Analyses.

Preliminary analyses were performed on all outcome responses (pre and post therapy) along with the WAI and RDI-R questionnaire responses using SPSS version 17. Pearson Correlations were conducted to identify significant bivariate relationships between criterion variables, which were the three outcome measures' (CORE, SI and PQ) post therapy scores, and predictor variables, which were the

three outcome measures' pre-therapy scores, the WAI and the RDI-R scores. Table 8.2 shows such correlations.

The correlations revealed that the predictor variable, RDI, significantly correlated with all three outcome measures' post-therapy scores. However, for the WAI, this was not the case, as it did not correlate significantly with any predictor or criterion variable. This indicates that the RDI-R has a moderate association with client post-therapy distress and suggests that presence of relational depth may influence outcome but that the WAI does not. However, these analyses do not control for the effect of pre-therapy nor do they explore their unique contribution to outcome.

Table 8.1. Clinical and reliable change as assessed by outcome measures

	Pre-therapy		**Post-therapy**				
	Mean	**SD**	**Mean**	**SD**	**Pre-post Effect size**	**Clinical Change**	**Reliable change**
CORE	1.68	.49	.96	.63	1.29	21(37)	28 (42)
SI	1.89	.48	2.74	.62	-1.55	20 (38)	24 (42)
PQ	5.03	.88	3.34	1.41	1.47	18 (40)	22 (42)
Average effect					1.44		

CORE: RCI = reduction of.5, clinical cut-off = 1.00 (clinical population a mean score of more than1).
SI: RCI is = increase of .6, clinical cut-off = 2.45 (clinical population a mean score of less than 2.45)
PQ: RCI = reduction of 1.43, clinical cut-off = 3.5 (clinical population a mean score of more than 3.5)

Table 8.2 Correlation matrix between predictor variables and criterion

	RDI-22	WAI	Pre-CORE	Post-CORE	Pre-SI	Post-SI	Pre-PQ
RDI-22							
WAI	.316						
Pre-CORE	.210	.204					
Post-CORE	-.381*	-.228	.265				
Pre-SI	.073	.118	.661**	.294			
Post-SI	-.487**	-.218	.155	.880**	.332*		
Pre-PQ	.036	.376	.487**	.014	.459**	.045	
Post-PQ	-.330*	-.061	.150	.622**	.262	.692**	.250

Note. * p< 0.05 . ** p< 0.01 Note. Sample sizes range from N=38 to N=43.

8.3.5 Regression Analyses.

Hierarchical multiple regression is used to determine how much variance in the criterion, dependent or outcome variable is explained by predictors (independent variables) when they are entered in a particular sequence' (Howitt & Cramer, 2011). Regression that focuses on finding an optimum set of predictors is stepwise and simultaneous regression. It does so by reducing the number of predictors but without reducing the coefficient significantly. Hierarchical regression is different as it focuses on the change in predictability associated with particular predictor variables entered later in the analysis over and above that contributed by predictor variables entered earlier in the analysis.

In order to examine the unique contribution of the RDI-R in the explanation of therapeutic outcome as assessed by the three outcome measures (CORE-OM, SI and PQ), a series of hierarchical multiple regression analyses were performed on the whole sample. Predictor variables were entered in three steps but in two different orders depending on the particular model. In Model 1, Pre-therapy was entered at Step 1, RDI-R was entered at Step 2 and WAI was entered at Step 3. This model tested for the unique contribution of working alliance on client improvement after pre-therapy and relational depth was controlled for. In Model 2, Pre-therapy was entered at Step 1, WAI at Step 2 and RDI-R at Step 3. This model tested for the unique contribution of relational depth on client improvement once pre-therapy and working alliance had been controlled for. Both Models 1 and 2 were each tested using all three outcome measures. This means that two hierarchical regression analyses were conducted for each measure totalling six regression analyses. A further two regression analyses were conducted which tested the mean outcome score

of all three outcome measures. This meant eight hierarchical regression analyses were conducted in all. Table 8.3 shows these regression analyses.

CORE Regression Analyses

Both models, Step 1 (Pre-therapy). The results of Step 1 for both models, where CORE-OM Pre-therapy score was entered into the equation, showed that the variance accounted for (R^2) in CORE-OM post-therapy score equalled .073, which is 7.3%. CORE–OM Pre-Therapy was not a significant predictor of CORE-OM post therapy score (p=.106).

CORE Model 1 (Step 2 = RDI, Step 3 = WAI): In step 2, the RDI-R was entered into the regression equation and the change displayed statistical significance (R^2 = .211; R^2 change =.138; significance of change = .020). The RDI-R alone accounted for 13.8% of the variance in outcome after pre-therapy client distress had been accounted for. In Step 3, the WAI was entered into the regression equation and the change did not display statistical significance (R^2 = .231; R^2 change =.020; significance of change = .355). The WAI alone accounted for 2% of the variance in outcome after pre-therapy and RDI had been accounted for. This indicates that, relational depth contributes significantly to a reduction in problems when pre-therapy state is controlled for. It also indicates that working alliance does not contribute to outcome when both pre-therapy and relational depth have been controlled for.

CORE Model 2 (Step 2 = WAI, Step 3 =RDI): In Step 2, the WAI was entered into the equation and the change did not display statistical significance (R^2 = .130; R^2 change =.057; significance of change = .144). The WAI alone accounted for 5.7% of

the variance in outcome after pre-therapy had been taken in account. In Step 3, the RDI was entered into the equation and the change displayed significance (R^2 = .231; R^2 change =.101; significance of change = .045). The RDI alone accounted for 10.1% of the variance in outcome after pre-therapy and WAI had been accounted for. This indicates that, when using the CORE-OM, working alliance does not contribute significantly to outcome when only pre-therapy is controlled for. It also indicates that relational depth significantly contributes to outcome when both pre-therapy and relational depth have been controlled for.

CORE-OM result summary. It would appear that 2% and 10.1% are the amounts of CORE-OM post-therapy variance that appear to be uniquely predicted by WAI and RDI respectively. This indicates that 3.7% is the amount of CORE post variance that is predicted by the overlap between RDI and WAI. This 3.7% has less of an impact on the 13.8% variance taken up by the RDI than the impact on the 5.7% variance taken up by the WAI. The RDI's unique variance takes up a much greater percentage of outcome than the WAI's. About 27% of the RDI that is accounted for by CORE-OM post-therapy (3.7 / 13.8), consists of the WAI. In contrast, 65% of the WAI accounted for by CORE OM (3.7 / 5.7) consists of the RDI.

The results in Table 8.3 (for the CORE-OM) indicate that, at least in this sample, after pre-therapy has been controlled for, relational depth contributes significantly to positive therapeutic outcome (as assessed by the CORE-OM) whether or not working alliance has been taken into consideration but that working alliance does not contribute significantly to outcome in the same way (after pre-therapy has been

controlled for and whether or not relational depth has been controlled for). These results also indicate that nearly two-thirds (65%) of the portion of working alliance that predicts outcome, is overlapped by relational depth. Conversely, the results also indicate that just over a quarter (27%) of the portion of relational depth that predicts outcome, is overlapped by working alliance. This would, in turn, imply that there is a greater amount of redundancy in working alliance than in relational depth when predicting therapeutic outcome as assessed by CORE-OM.

Table 8.3 Hierarchical Regression Analyses

Outcome measure Model	Step	B	SE	Beta	R^2	R^2 change	Sig. change
CORE	1. CORE Pre	.330	.199	.270	**.073**	.073	.106
Model 1	2. RDI	-.508	.208	-.376	**.211**	.138	.020
	3. WAI	-.153	.163	-.152	**.231**	.020	.355
Model 2	2. WAI	-.245	.164	-.244	**.130**	.057	.144
	3. RDI	-.452	.217	-.355	**.231**	.101	.045
Adjusted R^2 = .162; F(3, 33) = 3.313; p = .032							
SI	1. SI Pre	.439	.218	.331	**.110**	.110	.052
Model 1	2. RDI	-.861	.191	-.587	**.455**	.345	.000
	3. WAI	-.046	.141	-.046	**.457**	.002	.744
Model 2	2. WAI	-.220	.165	-.218	**.156**	.047	.192
	3. RDI	-.841	.203	-.574	**.457**	.300	.000
Adjusted R^2 = .404 ; F(3, 31) = 8.68; p = .000							
PQ	1. PQ Pre	.334	.253	.215	**.046**	.046	.195
Model 1	2. RDI	-1.084	.468	-.356	**.173**	.127	.026
	3. WAI	-.094	.402	-.042	**.174**	.001	.816
Model 2	2. WAI	-.373	.397	-.165	**.070**	.023	.354
	3. RDI	-1.045	.503	-.343	**.174**	.105	.045
Adjusted R^2 = .101; F(3,34) = 2.39; p = 086.							
All	1. All Pre	.344	.193	.285	**.081**	.081	.082
Model 1	2. RDI	-.745	.259	-.422	**.258**	.176	.007
	3. WAI	-.094	.207	-.072	**.262**	.005	.652
Model 2	2. WAI	-.269	.210	-.206	**.123**	.041	.207
	3. RDI	-.703	.278	-.398	**.262**	.140	.016
Adjusted R^2 = .197; F(3,34) = 4.02; p = .015							

Strathclyde Inventory (SI) Regression Analyses

Both models, Step 1: Pre-therapy. The results of Step 1 for both models, where SI Pre-therapy score was entered, indicated that the variance accounted for (R^2) with SI pre-therapy score equalled .110, which is about 11% of the variance in SI outcome. This was not a significant predictor.

SI Model 1. In step 2, the RDI-22 was entered into the regression equation and the change displayed statistical significance (R^2 = .455; R^2 change =.345; significance of change = p<.001). The RDI-22 alone accounted for 34.5% of the variance in outcome after pre-therapy had been accounted for. In Step 3, the WAI was entered into the regression equation and the change did not display statistical significance (R^2 = .457; R^2 change =.002; significance of change = .744). The WAI alone accounted for .2% of the variance in outcome after pre-therapy and RDI had been accounted for. This indicates that relational depth contributes significantly to outcome when pre-therapy is controlled for but that working alliance does not contribute to outcome significantly when both pre-therapy and relational depth have been accounted for.

SI Model 2. In Step 2, the WAI was entered into the equation and the change did not display statistical significance (R^2 = .156; R^2 change =.047; significance of change = .192). The WAI alone accounted for 4.7% of the variance in outcome as assessed by the SI when Pre-therapy is accounted for. In Step 3, the RDI was entered into the equations and the change displayed significance (R^2 = .457; R^2 change =.300; significance of change = p<.001). The RDI alone accounted for 30.1% of the variance in outcome after pre-therapy and WAI had been accounted for. This indicates that working alliance does not contribute significantly to outcome when

only pre-therapy (and not RDI) is controlled for. It also indicates that relational depth contributes significantly to positive therapeutic outcome after pre-therapy and working alliance have been taken into consideration.

SI Result Summary. Results indicate that .2% and 30.1% are the amounts of SI post-therapy variance that are uniquely predicted by the WAI and the RDI respectively. It also indicates that 4.5% is shared by WAI and RDI. This 4.5% has a considerable impact on the 4.7% taken up by the WAI but much less of an impact on the 34.5% of the RDI. The RDI's unique variance is considerably more than the WAI's unique variance on outcome as assessed by the SI. More specifically, 95% (4.5/4.7) of the portion of the WAI that predicts outcome, is taken up by the RDI, but only 15% (4.5/34.5 of the portion of the RDI that predicts outcome, is taken up by the WAI.

The above results (for the SI) indicate that for the present sample after pre-therapy has been controlled for, relational depth contributes significantly to positive therapeutic outcome (as assessed by the SI) whether or not working alliance has been taken into consideration but that working alliance does not contribute significantly to outcome in the same way (after pre-therapy has been controlled for and whether or not relational depth has been controlled for). The results also indicate that virtually all (95%) of the portion of working alliance that predicts outcome (as assessed by the SI), is explained by relational depth. In sharp contrast, only 15% of the portion of relational depth that predicts outcome is explained by working alliance. This in turn would imply that relational depth largely predicts most of the positive therapeutic

outcome that working alliance predicts but that working alliance predicts a very small proportion of the positive therapeutic outcome that relational depth predicts.

Personal Questionnaire Regression Analyses

Both models, Step 1. The results of Step 1 for both models, where PQ Pre-therapy score was entered, indicated that the variance accounted for (R^2) with PQ pre-therapy score equaled .046, which is about 4.6% of the variance in outcome and was not a significant predictor.

Model 1: In step 2, the RDI-R was entered into the regression equation and the change displayed statistical significance (R^2 = .173; R^2 change =.127; significance of change = .026). The RDI alone accounted for 12.7%% of the variance in outcome after pre-therapy had been accounted for. In Step 3, the WAI was entered into the regression equation and the change did not display statistical significance (R^2 = .174.; R^2 change =.001; significance of change = .816). The WAI alone accounted for .1% of the variance in outcome after pre-therapy and RDI had been accounted for.

This indicates that relational depth contributes significantly to outcome when pre-therapy is controlled for. It also indicates that working alliance does not contribute to outcome when both pre-therapy and relational depth have been accounted for.

Model 2: In Step 2, the WAI was entered into the equation and the change did not display statistical significance (R^2 = .070; R^2 change =.023; significance of change = .354). The WAI alone accounted for 2.3% of the variance in outcome. In Step 3, the RDI was entered into the equations and the change displayed significance (R^2 = .174; R^2 change =.105; significance of change = .045). The RDI alone accounted for

10.5% of the variance in outcome after pre-therapy and WAI had been accounted for. This indicates that working alliance does not contribute significantly to outcome when only pre-therapy (and not RDI) is controlled for. It also indicates that relational depth contributes significantly to positive therapeutic outcome after pre-therapy and working alliance have been taken into consideration.

PQ result summary. It would appear that .1% and 10.5% are the amounts of PQ post-therapy variance that appear to be uniquely predicted by WAI and RDI respectively. This indicates that 2.2% is the amount of PQ post-therapy variance that is predicted by the overlap between RDI and WAI. This 2.2% has less of an impact on the 12.7% variance taken up by the RDI than the impact on the 2.3% variance taken up by the WAI. The RDI's unique variance takes up a much greater percentage of PQ post-therapy outcome than the WAI's unique variance does. About 17% (2.2/12.7) of the RDI that is accounted for by PQ post-therapy consists of the WAI. In contrast, 96% (2.2/2.3) of the WAI accounted for by PQ-post therapy consists of the RDI.

The above results (for the PQ) indicate that, at least for this sample, after pre-therapy has been controlled for, relational depth contributes significantly to positive therapeutic outcome (as assessed by the PQ post-therapy outcome) whether or not working alliance has been taken into consideration but that working alliance does not contribute significantly to outcome in the same way (after pre-therapy has been controlled for and whether or not relational depth has been controlled for). These results also indicate that nearly all (96%) of the portion of working alliance that predicts outcome, is overlapped by relational depth. Conversely, the results also

indicate that about a sixth (17%) of the portion of relational depth that predicts outcome, is overlapped by working alliance. This would, in turn, imply that there is a greater amount of redundancy in working alliance than in relational depth when predicting therapeutic outcome as assessed by the PQ outcome measure.

Average Outcome Regression Analyses

Step 1, both models. In the results described here, the CORE, PQ and SI were totalled and mean average calculated. Therefore this variable will be referred to as *Average Outcome* from here on. The results of Step 1 for both models, where Average Outcome score was entered, indicated that the variance accounted for (R^2) with Average Outcome pre-therapy score equalled .081, which is about 8.1% of the variance in outcome and was not a significant predictor.

Model 1: In Step 2, the RDI was entered into the regression equation and the change displayed statistical significance (R^2 = .258; R^2 change =.176; significance of change = .007). The RDI-R alone accounted for 17.6% of the variance in Average Outcome after pre-therapy had been accounted for. In Step 3, the WAI was entered into the regression equation and the change did not display statistical significance (R^2 = .262; R^2 change =.005; significance of change = .652). The WAI alone accounted for .5% of the variance in Average Outcome after pre-therapy and RDI had been accounted for. This indicates that relational depth contributes significantly to outcome (Average Outcome) when only pre-therapy is controlled for. It also indicates that working alliance does not contribute to outcome when both pre-therapy and relational depth have been accounted for.

Model 2: In Step 2, the WAI was entered into the equation and the change did not display statistical significance (R^2 = .123; R^2 change =.041; significance of change = .207). The WAI alone accounted for 4.1% of the variance in Average Outcome when pre therapy had been controlled for. In Step 3, the RDI was entered into the equation and the change displayed significance (R^2 = .262; R^2 change =.139; significance of change = .016). The RDI alone accounted for 13.9% of the variance in outcome after pre-therapy and WAI had been accounted for. This indicates that working alliance does not contribute significantly to outcome when only pre-therapy is controlled for. It also indicates that relational depth contributes significantly to positive therapeutic outcome after pre-therapy and working alliance have been taken into consideration.

Average Outcome Results Summary. It would appear that .05% and 14% are the amounts of Overall post-therapy variance that appear to be uniquely predicted by WAI and RDI respectively. This indicates that 3.6% is the amount of Overall post-therapy variance that is predicted by the overlap between RDI and WAI. This 3.6% has less of an impact on the 17.6% variance taken up by the RDI than the impact on the 4.1% variance taken up by the WAI. The RDI's unique variance takes up a much greater percentage of Overall post-therapy outcome than the WAI's unique variance does. About 20% (3.6/17.6) of the RDI that is accounted for by Overall post-therapy consists of the WAI. In contrast, 88% (3.6/4.1) of the WAI accounted for by Overall-post therapy consists of the RDI.

The above results indicate that, for the present sample, after pre-therapy has been controlled for, relational depth contributes significantly to positive therapeutic

outcome (as assessed by an average of all the measures) whether or not working alliance has been taken into consideration; but working alliance does not contribute significantly to outcome in the same way (after pre-therapy has been controlled for and whether or not relational depth has been controlled for). These results also indicate that a large proportion working alliance that predicts outcome is overlapped by relational depth. Conversely, the results also indicate that about 20% of the portion of relational depth that predicts outcome is overlapped by working alliance. This would, in turn, imply that there is a greater amount of redundancy in working alliance than in relational depth when predicting therapeutic outcome as assessed by the post-therapy outcome measures as a whole.

8.3.6 Social Anxiety versus Practice-Based Clients

To explore more closely the difference between the Social Anxiety sample and the Practice-Based sample, comparison analyses were conducted for all predictor variables. There was a statistically significant difference (p. =.02) for Pre-therapy PQ scores where clients with social anxiety scored more highly (mean = 5.48) than the practiced-based clients (mean = 4.85). This indicates that SA clients on average experience higher levels of specific problem-related distress. No statistically significant differences were found between these two groups for any of the other predictor variables. Table 8.4 shows means, and effects sizes for all predictor variables for Social Anxiety clients and Practice-Based clients.

Separate hierarchical multiple regression analyses were carried out for each of these populations, in each of the outcome measures. Table 8.5 shows results of such separate multiple regression analyses.

Table 8.4. Mean RDI-R scores for SA vs PB clients with effect sizes

	SA *N*= 13		PB *N*=21		Effect Size
	Mean	SD	Mean	SD	
RDI	2.81	0.33	2.93	0.51	-0.29
WAI	3.99	0.51	3.85	0.67	0.24
CORE Pre	1.61	0.54	1.72	0.47	-0.22
CORE Post	1.18	0.68	.91	0.62	0.41
SI Pre	2.10	0.44	2.16	0.54	-0.12
SI Post	1.54	0.50	1.17	0.67	0.63
PQ Pre	5.48	0.83	4.85	0.83	0.76*
PQ Post	3.76	1.36	3.22	1.35	0.40

Results of the separate hierarchical regression analyses showed that, for Social Anxiety clients, when pre-therapy was controlled for, the RDI was not a significant predictor of positive therapeutic outcome, whether the WAI was controlled for or not. In addition, the WAI was not a significant predictor of positive therapeutic outcome, whether the RDI was controlled for or not. These results were consistent across all outcome measures. It is worth noting that although statistical power was quite low for the SA analyses, the R Squared change values for the RDI were 2% of variance across the board, suggesting that the most likely interpretation is that there is no relationship between RDI and outcome for the SA clients rather than low power being the problem. In sharp contrast, for PB clients, when pre-therapy was controlled for (in each outcome measure), the RDI was always a significant predictor of positive therapeutic outcome whether WAI was controlled for or not. Conversely, the WAI was never a significant predictor of positive therapeutic outcome whether the RDI was controlled for or not.

Table 8.5: Separate Multi-regression Analyses for SA and PB clients.

	Step	B	SE	Beta	R^2	R^2 change	Sig. change
	Social Anxiety Clients						
	1: CORE Pre	.380	.318	.327	.107	.107	.254
Model 1	2: RDI	.221	.548	.116	.013	.013	.694
	3: WAI	-.524	.392	-.406	.134	.134	.211
Model 2	2: WAI	-.460	.373	-.357	.215	.109	.243
	3: RDI	.388	.544	.204	.253	.038	.492
CORE Adjusted R^2 = .029; F(3,10) = 1.131; p = .383							
	1. SI Pre	.550	.250	.536	.287	.287	.048
Model 1	2: RDI	.045	.374	.033	.288	.001	.905
	3: WAI	-.134	.260	-.143	.307	.018	.618
Model 2	2: WAI	-.121	.242	-.129	.303	.016	.628
	3: RDI	.090	.397	.065	.307	.004	.825
SI Adjusted R^2 = .099; F(3,10) = 1.475; p = .280							
	1. PQ Pre	.011	.454	.007	.000	.000	.981
Model 1	2: RDI	-.221	1.186	-.054	.003	.003	.855
	3: WAI	-.618	1.132	-.232	.029	.026	.596
Model 2	2: WAI	-.552	.939	-.207	.028	.028	.567
	3: RDI	.165	1.412	.040	.029	.001	.909
PQ Adjusted R^2 = -.235; F(3,11) = .111; p = .952							
	1: Average Pre	.286	.252	.300	.090	.090	.277
Model 1	2: RDI	-.018	.639	.008	.090	.000	.978
	3: WAI	-.412	.430	-.285	.160	.070	.358
Model 2	2: WAI	-.363	.385	-.251	.153	.063	.365
	3: RDI	.219	.687	.099	.160	.008	.755
AVERAGE Adjusted R^2 = -.069; F(3,11) = .701 ; p = .571							
	Practice-Based Clients						
	1. CORE Pre	.341	.255	.280	.078	.078	.196
Model 1	2: RDI	-.637	.203	-.556	.382	.304	.005
	3: WAI	-.118	.167	-.134	.398	.016	.489
Model 2	2: WAI	-.264	.181	-.301	.167	.088	.161
	3: RDI	-.587	.217	-.512	.398	.231	.014
CORE - Adjusted R^2 = .303; F(3,19) = 4.184; p = .020							
	1. SI Pre	.422	.302	.306	.093	.093	.178
Model 1	2: RDI	-1.056	.217	-.722	.609	.516	.000
	3: WAI	-.084	.166	-.084	.615	.006	.619
Model 2	2: WAI	-.331	.218	-.332	.196	.103	.147
	3: RDI	-1.014	.236	-.694	.615	.419	.000
SI - Adjusted R^2 = .547; F(3,17) = 9.051 ; p = .001							
	1. PQ Pre	.378	.337	.238	.057	.057	.274
Model 1	2: RDI	-1.298	.499	-.493	.295	.238	.017
	3: WAI	.033	.423	.016	.295	.000	.939
Model 2	2: WAI	-.290	.448	-.144	.076	.019	.525
	3: RDI	-1.311	.539	-.498	.295	.219	.025
PQ - Adjusted R^2 = .184; F(3,19) = 2.653 ; p = .078							
	1: Average Pre	.277	.300	.197	.039	.039	.366
Model 1	2: RDI	-.879	.278	-.566	.359	.320	.005
	3: WAI	-.065	.239	-.055	.361	.002	.788
Model 2	2: WAI	-.290	.262	-.244	.095	.056	.280
	3: RDI	-.851	.302	-.549	.361	.267	.011
AVERAGE - Adjusted R^2 = .260; F(3,19) = 3.58; p = .033							

Summary of Outcome Regression Analyses for PB v SA Clients

All the regression analyses conducted for the PB sample and the SA sample are summarised here rather than every particular analysis described in detail (as has been done with the whole sample above). This is to save repetition and over-documenting results which can be inspected in Table 8.5 especially since the PB v SA regression analyses were not the main focus of this particular piece of research.

Average Outcome Regression Analysis for Practice-Based Clients Only

Average Outcome Model 1 and *Model 2 for Practice-Based Clients Only*. The Average Outcome refers to the three outcome measures (CORE, PQ and SI) which were totalled and mean averages calculated. The results of Step 1 for both models, where Average Outcome score (for Practice-Based clients only) was entered, indicated that the variance accounted for (R^2) with Average pre-therapy score equalled .039, which is about 3.9% of the variance in outcome and was not a significant predictor (p = .366)

Model 1: In step 2, the RDI was entered into the regression equation and the change displayed statistical significance (R^2 = .359; R^2 change =.320; significance of change = .005). The RDI-R alone accounted for 32% of the variance in outcome after pre-therapy had been accounted for. In Step 3, the WAI was entered into the regression equation and the change did not display statistical significance (R^2 = .361; R^2 change =.002; significance of change = .788). The WAI alone accounted for .2% of the variance in outcome after pre-therapy and RDI had been accounted for.

This indicates that relational depth contributes significantly to outcome when only pre-therapy is controlled for. It also indicates that working alliance does not contribute to outcome when both pre-therapy and relational depth have been accounted for.

Model 2: In Step 2, the WAI was entered into the equation and the change did not display statistical significance (R^2 = .095; R^2 change =.056; significance of change = .280). The WAI alone accounted for 5.6% of the variance in outcome when pre therapy had been controlled for. In Step 3, the RDI was entered into the equation and the change displayed significance (R^2 = .361; R^2 change =.267; significance of change = .011). This indicates that working alliance does not contribute significantly to outcome when only pre-therapy is controlled for. It also indicates that relational depth contributes significantly to positive therapeutic outcome after pre-therapy and working alliance have been taken into consideration.

Average Outcome Results Summary. It would appear that .2% and 26.7% are the amounts of Average post-therapy variance that appear to be uniquely predicted by WAI and RDI respectively. This indicates that 5.3% is the amount of Overall post-therapy variance that is predicted by the overlap between RDI and WAI. This has less of an impact on the 32% variance taken up by the RDI than the impact on the 5.6% variance taken up by the WAI. The RDI's unique variance takes up a much greater percentage of Average post-therapy outcome than the WAI's unique variance does. About 17% (5.3/32) of the RDI that is accounted for by Average post-therapy

is explained by the WAI. In contrast, 95% (5.3/5.6) of the WAI accounted for by Average-post therapy consists of the RDI.

Indications of Average Outcome result for PB clients. The above results indicate that, for the PB sample, after pre-therapy has been controlled for, relational depth contributes significantly to positive therapeutic outcome (as assessed by an average of all the measures) whether or not working alliance has been taken into consideration. However, working alliance does not contribute significantly to outcome in the same way (after pre-therapy has been controlled for and whether or not relational depth has been controlled for). This is the pattern which ever outcome measure is being used and the Average Outcome result is indicative of this.

These results also indicate that a large proportion of working alliance that predicts outcome, is overlapped by relational depth. Conversely, only a small proportion of relational depth (that predicts outcome) is overlapped by working alliance. This would, in turn, imply that there is a greater amount of redundancy in working alliance than in relational depth when predicting therapeutic outcome as assessed by the post-therapy outcome measures as a whole.

Average Outcome Regression Analyses for Social Anxiety Clients

The regression analyses results for Social Anxiety clients are not documented in detail here simply because none of such results yielded any statistical significance. In both Models 1 and 2 there were no significant changes in any of the steps taken in the regression analyses. This can be seen in Table 8.5. These results would suggest

that although Social Anxiety clients may experience moments of relational depth in therapy, such clients cannot make use of the experience to facilitate improvement.

8.4 Chapter Summary

In using a more reliable RDI-R, Study 3 provided evidence that a significant event that has a presence of relational depth is related to overall therapeutic improvement. Even though there are issues with this study already discussed in Study 3's discussion section (i.e. timing of administration of WAIs and RDIs in that both were completed at the same stage of therapy as the outcome measures were), it still is arguably strong evidence that a moment of relational depth can positively affect improvement in therapy. These results are further discussed in the main discussion section.

Chapter 9: Main Discussion

9.1 Introduction to Chapter

This last chapter in this PhD research project is designed to bring all three research studies together and discuss their findings as a whole in light of the research project's main objectives. I first summarise all the findings together and then go on to discuss why the findings are important. I further go on to discuss the implications for therapy practice and then relate these finding to other studies which focus on relational depth. I give suggestions for further research and lastly give a conclusion.

9.2 Major Findings of Project

There were essentially two keys objectives of this whole research project: to test the Relational Depth Inventory (RDI) and refine the RDI before using it to investigate relational depth and client improvement during therapy.

Each of the three studies in this project aimed to provide evidence of, and to improve, the reliability and validity of the Relational Depth Inventory. Once reliability and content-validity evidence had been established, an additional objective was to explore the relation between relational depth and outcome. This would serve to provide predictive validity evidence. Across all three studies, the main objective has been met in that the RDI has been developed and improved with each study being a step in its development. In the final study the additional objective was met in that the RDI helped provide evidence that relational depth is predictive of client improvement.

9.3 Meaning of Findings and Why They are Important

The current project and the findings are important because they add to the growing body of research into relational depth. Essentially, all three studies provide add to the growing evidence that relational depth exists. More importantly, with respect to the research objectives here, the three studies evidence that the Relational Depth Inventory shows considerable reliability and validity evidence and that relational depth is related to client improvement.

In Study 1, factor analytic results would suggest that relational depth is composed of several possible elements, especially those associated with deep relational experiences such as love, connectedness and respect. In addition, experiences of transcendence appear to constitute a specific component of relational depth, labelled by informants as "spiritual" and "magical". This implies that relational depth can include experiences that go beyond everyday therapeutic encounters.

However, a limitation of the factor analyses in Study 1 is that none of these analyses gave any indication as to how unidimensional the RDI is. More importantly they do not reveal any secondary dimensions. Furthermore, none of the Classical Test Theory methods of Study 1 were able to group persons as to how persons may be grouped in relation to presence of relational depth nor did they show difficulty of items. In sharp contrast, The Rasch Analyses in Study 2 were able to do just this.

In Study 2, in the Rasch analysis, although there was excellent reliability evidence, there was evidence that not all 24 items formed a unidimensional measure. Items had to be removed in order to provide optimum reliability for persons and to improve

infit and outfit statistics. Items measuring 'feeling like going beyond ordinary limits' and 'feeling like time had stopped' weakened reliability and were thus removed in order to improve reliability. This would indicate that such experiences as going beyond the ordinary and a feeling of time stopping may not have anything to with the dimension of relational depth. Another explanation for why these items weakened reliability could be that the language may have been seen to be a little jargon-like or that the meaning could have been ambiguous; for example, the phrase 'going beyond ordinary limits' may have been taken to mean a number of things including the breaking of professional and/or personal boundaries, which would theoretically not be part of relational depth.

The item hierarchy (in Study 2) indicated that there are 11 distinct groups (strata) of items on the RDI, with items in each strata containing items of roughly equal difficulty. Items easiest to endorse assessed those experiences that concerned the client's experience of the therapist indicating that this is a basic presence of relational depth or a foundational presence. The moderately difficult to endorse items assessed aspects of the relationship, while the harder to endorse items included items that assessed the client understanding the therapist as well as spiritual experiences. This would imply that spiritual experiences are much rarer aspects of relational depth.

At the time of writing, there is no specific theory on relational depth that looks at which relational depth experiences may be stronger, weaker or more difficult to experience. However, there is previous research is which supports the current research findings. In a qualitative study, Knox (2008) found that out of three aspects

of clients' experiences of a moment of relational depth (self, therapist and relationship) it was the relationship aspect where clients reported to have 'deeper' feelings. Clients also reported that they felt the experience of the relationship was on 'a different dimension' (compared to experience of self and therapist) and that experiences of feeling spiritual, mystical and mutual understanding accompanied this. Such findings are consistent with the item hierarchy in the research presented here, where items assessing spiritual experiences and the client understanding the therapist were most difficult to endorse.

In Study 3, The major findings were that RDI-R scores significantly predicted outcome scores as assessed by three different outcome measures. Furthermore, such scores were predicted when pre therapy outcome scores and WAI scores were controlled for. This is therefore the first quantitative evidence that outcomes in therapy are associated with moments of in-depth connection – over and above the alliance. This indicates that the stronger the presence is of relational depth in a helpful event, the more positive is the therapeutic outcome or improvement. In spite of the strong overlap between WAI and RDI, the WAI shows little or no relationship to outcome. Furthermore, what relationship to outcome it shows, is shared with the RDI. This suggests that the presence of relational depth during a helpful event is predicting something different in improvement than working alliance is. In other words this result is indicative of there being something which leads to improvement which working alliance does not predict but that relational depth does. It is important to note, however, that these results were not consistent across all of the sample. For the clients in the study who suffer with social anxiety, relational depth presence in a

helpful event did not predict improvement by any of the outcome measures. This is despite there being no difference in RDI-R scores between social anxiety clients and non-social anxiety clients. This would suggest that social anxiety clients are able to experience relational depth within therapy but are not able to use that experience to help them outside of therapy.

The strong association between RDI Index and WAI-SR, indicates either that Relational Depth is more likely to occur in the presence of a good therapeutic alliance or that there is substantial conceptual overlap between relational depth and therapeutic alliance. Nevertheless, this high correlation does not preclude the idea that relational depth may be a higher or upward extension of working alliance, tapping a different region of the relational continuum. However, further research is needed to explore the nature of this overlap and to test whether this is the case.

9.4 Relating Findings to Other Studies

In the first study (Study 1), items assessing respect, intimacy, mutuality, feeling at one with the other, and meeting of minds strongly characterized relational depth presence in the discriminant function analyses, thus supporting Mearns and Cooper's (2005) account. The item 'being in the moment' also correlated with relational depth presence and suggests that the experience of relational depth is more a short-lived event than an enduring experience. The items 'still atmosphere' and 'timeless atmosphere' were also associated strongly with relational depth presence and would appear to be characteristic of particular moments rather than general relational processes. This finding also supports Stern's (2004) idea of moments of meeting in

which each party is aware of what the other is experiencing. This is consistent with the notion that therapeutic relationship is developed or highlighted via momentary intersubjective events.

Being available and being real were also important elements of relational depth, consistent with previous research, notably Knox (2008), who found that clients experienced their therapists as providing psychological holding and being real during times of relational depth. The results of the discriminant function analyses of the RDI also showed that experiences of connection and love are highly characteristic of relational depth. The word "connection" was frequently used by Mearns and Cooper (2005) in describing relational depth and this result is consistent with their theory. However, these authors make no mention of the experience of love and this result was largely unexpected. Nevertheless, Mearns and Thorne (2000) noted that training to be a person-centred counsellor involves a process of freeing one's 'loving self', while Keys (2007) has argued for the central role of the experience of love in therapy, proposing four different dimensions of therapeutic love: unconditional positive regard, contact/perception, empathy, and congruence. In putting this forward she indicates that the emergence of love in therapy functions as a healing quality in the therapeutic relationship. The result reported here (where the item 'love' correlates with relational depth presence), is thus quite consistent with the person-centred view of relational depth as being a combination of all six of Rogers' (1957) necessary and sufficient conditions.

In Study 1, the RDI was a symmetrical measure of therapeutic relationship in significant events assessing both clients' and therapists' presence of relational depth. Therefore it is interesting to compare client and therapist views obtained with it. For example, significant events rated as indicating a presence of relational depth were no more likely to occur for clients than therapists. However, therapists did score higher on the RDI Index than clients, indicating that it might be more sensitive to role differences more than the content analysis measure (RD Presence). Moreover, it is evident that clients and therapists in the study experienced relational depth differently on the five RDI factors obtained for the whole sample. Given the different roles of each party in the therapeutic relationship, therapists focused on empathising and respecting the other (Factor 1), while clients experienced receiving of empathy from the other (Factor 5). This is hardly surprising. In addition, in the separate factor analyses, Item 1 ("liberated") was part of one of the main factors for clients but not therapists. This is consistent with McMillan and McLeod (2006), who found that 'letting go' was an aspect of the client's experience of relational depth.

The separate factor analyses for clients and therapists would seem to confirm that there is a difference of experience between these two groups. This would be particularly relevant with regard to the empathy items where for clients 'other empathic towards me' was a high loading item and 'empathy for other' much lower. The opposite was true for therapists. However, results of the separate factor analyses need to be treated with caution due to the smaller sample size in each analysis. Further research comparing client and therapist experiences of relational depth would thus be useful as would the development of separate RDIs (for clients and therapists).

Experiences of being scared and vulnerable, reported by some participations in the prior qualitative studies used for item generation, were not associated with the other aspects in the reliability and factor analyses, suggesting that such experiences are not part of the main relational depth construct. Relational fear/vulnerability may be an important but separate phenomenon found in a subset of clients and therapists. Mearns and Cooper (2005), for example, talks about the fear in the client when they may begin to engage at relational depth; in other words the client may see the offer of 'an in-depth understanding as both deeply attractive and deeply terrifying' (p.52). Mearns (2009) also acknowledges the fear that may be felt by a client when they have been damaged in relationships previously and in engaging at relational depth, the client may be going where they are too scared to go. The presence of this 'scared/vulnerable' factor as an independent aspect of relational depth also parallels the identification of a roughness/distress dimension of therapy process alongside depth/value/effectiveness in research by Orlinsky & Howard (1977; 1986) and Stiles and Snow (1984). In addition, Process-Experiential/Emotion-Focused Therapy incorporates a strongly relational task referred to as "Empathic affirmation at vulnerability" in which a deep relational connection is offered to a client who is in an intensely vulnerable place (Elliott, Watson, Goldman & Greenberg, 2004). With this in mind further research is recommended on the phenomenon of relational fear/vulnerability, e.g., with socially anxious clients.

In the Rasch analysis study (Study 2), more specifically, in testing the RDI for its unidimensionality, the two contrasting dimensions conform with the Smith-

Benjamin's (1996) Structural Analysis of Social Behaviour (SASB) model of interpersonal and intrapsychic events. This model is divided into three underlying dimensions; a) focus (on other, on self with other, and on self with self, b) affiliation (love vs hate), and c) interdependence (enmeshment vs differentation). The focus dimension's aspect of focus on the other is usually concerned with a parent-like focus and usually pertains to the therapist's focus on the client. In contrast, the focus on self with other, and self with self, is often used by the SASB model to apply to therapist or client. According to this model, for each of the dimensions, there are corresponding experiences and/or behaviours. For example, the model postulates that if self-focus, in relation to the other (the therapist), is characterised by 'enthusiastic showing' then the corresponding self-focus in relation to self may be a feeling of being 'pleased with self.' The model postulates that there will also be corresponding patterns of affiliation and interdependence, which in the examples given, would be self-affirmation.

With regard to the last piece of research in this project (Study 3), there are a number of studies which relate to this process-outcome study. Firstly, as mentioned in the literature review, in a quantitative study, Leung (2009) found that ratings of the extent to which relational depth had a high enduring effect were high. Knox (2008), in a qualitative study, found that the majority of participants said they felt that relational depth had a positive effect on therapeutic process. The current research, therefore, is compatible with this previous research in that it evidences that relational depth contributes to therapeutic improvement. More importantly, the research (Study 3) would suggest that because the RDI-R predicts therapy improvement, it evidences

predictive validity and therefore this study provides construct-related evidence for the RDI-R.

9.5 Implications for Practice

The results of this study have implications for therapists in that they would suggest that therapists may need to be open to meeting a client at a depth of intimacy in order to facilitate relational depth and potentially positive outcome. What is most surprising about the current study is that it would appear that one helpful event appears to play a significant role in positive outcome whereas the ongoing quality of the relationship of working alliance did not. This emphasises the role of significant events in the process of therapy and therefore suggests that the ongoing quality of the relationship, although important, may not play as large a role in outcome compared to shorter lived experiences that are important moments in therapy.

For therapy practice the evidence here suggests that therapists may be able to help clients by being able to engage with clients at a deep level. Therefore those therapists who prefer a 'professional distance' with their clients may not be able or not want to engage at depth. If the therapist does not or cannot engage at such depth then the research evidence here would suggest that this may be not as helpful to the therapy improvement or outcome, at least in Person-Centred Therapy. In other words, disconnection between therapist and client, due to one or both parties not engaging may have a negative affect on improvement or outcome. However, it is important to mention at this point that, paradoxically, an attempt by the therapist to

encourage or lead a client into relating at depth would be futile in that it would, theoretically, have the effect of distancing the client. The state of relational depth, theoretically, requires that both client and therapist each possess the high levels of unconditional positive regard, empathy and congruence. Further, an attempt by either party to direct or lead the other to experience relational depth actually requires a lack of these core conditions; trying to lead or direct somebody into any experience requires a distinct lack of unconditionality and empathy at least. What the current research does imply is that unconditionality, empathy and genuineness characterise relational depth and this is particularly evident from the results of Studies 1 and 2 particularly.

In terms of clinical implications of the relational depth concept, to this I will only add it is my view that therapists practicing in a variety theoretical orientations may find it important to be able to work at higher levels of relatedness with their clients. For these therapists, results suggest that they may need to be open to experiencing intimacy, love and possibly even transcendence. "Love" here is not meant in any romantic sense but rather as a therapeutic, humanitarian love (cf. Greek *philia* or perhaps even *agape*). Such experiences may not be possible for therapists who would wish to keep a professional distance, but nevertheless open up possibilities for genuine transformation via the therapeutic relationship.

9.6 Alternative Explanations of Findings

Alternative explanations for the results in all three studies are that it is possible that the presence of relational depth was incidental to the helpful or important event. In other words, there may have been an enduring presence of relational depth

throughout the therapy and not simply a momentary experience of it. If this were the case, this would mean that the RDI-R was actually assessing the enduring sense of contact rather than a momentary sense of contact. This might suggest that what was actually being measured was the ongoing deep therapeutic relationship rather than a moment of relational depth. In other words relational depth was a continuous presence and was there present throughout therapy including the important and helpful events..

In Study 3 an alternative explanation as to why the WAI scores were not significantly associated with the RDI-R scores or outcome measure scores could be due to the fact that although they were completed at the same stage of therapy, they were completed at a different time and possibly in a different place. Outcome packs were often sent out to clients and so it is possible that both RDI-R and outcome measures were completed in the client's own home, whereas the WAIs were generally completed at the research clinic. It is quite possible that this difference in time and environment could affect the difference in scores. However, perhaps the most likely explanation as to why WAI scores did not predict outcome is due to the particular WAIs used; the WAI scores which were analysed were completed at the same stage of therapy as the outcome measures. Horvath (1994) states that 'research results indicate the relation between session level outcome and working alliance measured at the same time is rather weak. On the other hand working alliance seems to be able to predict outcome' (p. 118). In other words, the WAIs scores used in the current study do not predict outcome because they were completed at the same stage of therapy that the outcome measures were completed. In a study by Watson and

Geller (2005) outcome measures were administered in subsequent sessions to WAIs. Therefore WAIs were always completed at least a week before outcome measures were. Watson and Geller (2005) found that WAIs predicted outcome in 3 out of 4 outcome measures.

9.7 Limitations

Study 1

In Study 1, There are several limitations with this study. First, many of the participants who responded as clients were likely to have been therapists drawing on their experiences as clients; this is largely because emails were sent to various directories of therapists. Clients who are also therapists are likely to have been trained to value the role of the therapeutic relationship, and to have developed more nuanced perceptions of therapy than clients who are not also therapists. While this is likely to make them more sensitive observers of their therapy experiences, it also means that they may represent a somewhat different population than nontherapist clients. Nevertheless, since therapists tend to be high utilizers of therapy (Orlinsky & Rønnestad, 2005), a sample biased toward such clients is not entirely unrepresentative. Research in progress with non-therapist client samples should throw light on this issue.

Second, the title of the questionnaire (In Study 1) explicitly pointed to the relationship between client and therapist. This may have elicited relationally-oriented therapists and significant event descriptions, thus artificially increasing the

proportion of relational depth significant events. Further research using other methods of eliciting significant descriptions, such as the more neutrally-worded Helpful Aspects of Therapy Form (Llewelyn, 1988), is needed to test whether these estimates of the rates of relational depth events are accurate.

Third, the authors made no attempt to collect data on the modality of the therapy and this may present difficulty in interpreting the results regarding the prevalence of relational depth events.

Fourth, participants were asked to recall an event that could have been months or years in the past. The reconstructive nature of memory may suggest that what was recalled concerned the characteristics of a more enduring relationship rather than the event. Although this is a possibility, the RDI is designed to elicit episodic, autobiographical memory rather than general, schematic memory, by asking participants to think about an event that 'stands out in your mind' and also to describe how they felt 'at that moment'.

Fifth, the online administration of the questionnaire may have excluded potential participants who were not computer literate or not comfortable with technology. In addition, over-zealous spam filters may have prevented some participants from receiving recruitment emails and links in the first place.

Sixth, dichotomizing RD Presence content ratings at the scale midpoint in order to designate RD events was problematic from a measurement point of view, because of

loss of information, unreliability leading to misclassification of events, and the unimodal (rather than bimodal) nature of the distribution.

It is worth noting that even for the events rated as containing relational depth, it is not known that relational depth was the factor responsible for their having been experienced as significant by participants. In other words, what is being called relational depth may have only been incidental to these events.

Study 2

The main limitations to Study 2 are thought to be the fact that a large part of the sample, the web-based participants, were responding retrospectively in that they could have been describing an event which happened many years ago. Therefore participant's memory may not have been completely reliable.

Another limitation may have been the low number of clients in RDI and WAI study and the non-independent data. This could have caused unreliable results.

In addition a criticism could be that the study has not provided divergent-related evidence in that there was no study or research in place where the RDI could have tested to explore whether it was unrelated to a concept that it should be unrelated to.

Study 3

The main limitations of Study 3 are probably due to the timing of the completion of WAIs. They were completed at the same stage of therapy as the outcome measures and therefore any predictive quality which the WAI has, was arguably not assessed. In Study 3 it may have been more productive to have used WAIs which had been

completed a few sessions before (e.g. where the outcome measure was completed at session 10, the WAI completed at session 5 could have been used). This is a little different for the RDI-R because clients here were asked about an event that had previously happened. Therefore, RDI-R can be completed at the same time as outcome measures simply because the event that is being described has already happened before any outcome measure is being completed. However, whilst this provides evidence that a relational depth event might predict therapeutic improvement, there is not evidence that the RDI-R has a predictive quality simply because RDI-Rs and outcome measures were completed at the same time. A further limitation could have been the small sample sizes of social anxiety clients, which could have contributed to a less powerful statistical analysis.

9.7 Suggestions for Further Research

With regard to research, the current project and findings provides researchers with numerous possibilities. In particular, this project provides the research community with a measure designed to assess moments of deep connection and therefore provides a valuable research tool that assesses aspects of the client-therapist relationship. The RDI-R also provides a new kind of tool to examine significant events in therapy.

Therefore, the RDI-R is able to facilitate studies in two main areas of research.

It is recommended that first the RDI-R have minor adjustments in that items are added that are designed to assess stronger presence of relational depth. This is due to Study 2's findings which suggested that there were persons not sufficiently targetted

by the RDI-R, who experienced stronger presence of relational depth than the RDI-R assesses. It is also recommended that the item 'going beyond my ordinary limits' be reworded due to some respondents saying they felt it referred to going beyond personal boundaries. At the time of writing these revisions are taking place.

Once the RDI-R has its added items reworded, and in light of Study 3's limitations, it is recommended that a process-outcome study take place where WAIs and RDI-Rs are administered at the same time in at least one session before outcome measures are administered in order to more accurately assess the relationship between the WAI and the RDI-R and to assess any predictive quality the RDI-R may have at predicting outcome and/or improvement. In addition, in order to test for convergent-related validity evidence, it is recommended that the RDI-R be administered with the BLRI so that the mutual experiences of empathy, unconditional positive regard and congruence can be assessed alongside the RDI. This is based on the fact that the RDI is theoretically assessing mutual levels of these conditions. In order to check for discriminant-related validity evidence it is suggested that a measure be designed to assess disconnection be created and then administered with the RDI-R. Such a dis-connection measure would be designed to explore the antithesis of relational depth. This, paradoxically, would still be researching relational depth, albeit indirectly, in that it would focus on how disconnection and absence of relational depth are associated with outcome and/or improvement.

Regarding Study 3, future research could improve on this study by repeating it and by administering the WAI and RDI-R as suggested above. In addition, a follow-up

question could be ‘do some types of significant events elicit relational depth presence more than others?’ This could be done by conducting a qualitative analysis of significant events where, if there emerges different groupings of significant events, such groupings’ RDI-R scores could be compared. This could help to identify which kinds of significant events are more likely to elicit the presence of relational depth. This could actually be done with the current data.

Another suggestion for further research would be to simply have a larger sample size especially where social anxiety clients are concerned. A total of 15 social anxiety clients is arguably not enough to conduct multiple regression as has been done here. Also concerning social anxiety, it would be productive if further research were conducted in order to explore why outcome for these clients was not predicted by either the RDI-R or the WAI. Qualitative analysis is suggested here as this may directly address the question as to why social anxiety clients do not seem to benefit from relational depth when most other clients do. For example analyzing the event descriptions in SA and PB on RDI and HAT to see if there is a difference in frequency of RD events.

The two research instruments introduced in Study 1, the RD content analysis method and the RDI, offer numerous possibilities for further research, for example, using more naturalistic samples of clients and more fine-grained psychometric analyses (e.g., Item Response Theory). Among other things, this would make it possible to test the hypothesis that the RDI offers an upward extension to instruments like the WAI. It will also be important to study convergent and predictive validity of the

RDI in a new prospective sample of clients, correlating it with other measures of relationship (e.g., Working Alliance Inventory) or session depth (i.e., the Session Evaluation Questionnaire), and therapy outcome. Other future uses of RD content analysis ratings and RDI include (a) prospective, longitudinal studies in which clients are asked to report on the occurrence of relational depth at regular intervals; (b) cluster analyses of different types of relational depth events, using client descriptions of important events or patterns of RDI item endorsement; and (c) use of RD content analysis or RDI to identify relational depth events for detailed analyses using methods such as Comprehensive Process Analysis (Elliott et al., 1994). The RDI could also be used to investigate associations between relational depth and the depth subscale in the Session Evaluation Questionnaire (Stiles, Gordon and Lani, 2002).

9.8 Conclusion

The research objectives throughout this whole PhD project have been met in that the RDI has been tested for reliability and validity using both tried and tested methods (Classical Test Theory) as well as using a more revolutionary method (The Rasch Model). Using both these methods served to test and improve the RDI. Once improvement of the measure had taken place, the revised Relational Depth Inventory (the RDI-R) was used to investigate relational depth and outcome. This therefore concludes this project.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement Education. (1985). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association

American Educational Research Association, American Psychological Association, & National Council on Measurement Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Andrich, D. (1988). *Rasch models for measurement*. Beverly Hills: Sage Publications

Azad, K. (2009). *Better explained: learn right not rote: an intuitive guide to exponential functions & e*. Retrieved on 22 October 2009 from http://betterexplained.com/articles/an-intuitive-guide-to-exponential-functions-e/

Barkham, M., Evans, C., Margison, F., McGrath, G., Mellor-Clark, J., Milne, D. & Connell, J. (1998). The rationale for developing and implementing core outcome batteries for routine use in service settings and psychotherapy outcome research. *Journal of Mental Health, 7,* 35-47.

Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., Benson, L., Connell, J., Audin, K. & McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE-OM: Towards practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology, 69,* 184-196.

Barkham, M. & Mellor-Clark, J. (2003). Bridging evidence-based practice and practice-based evidence: developing a rigorous and relevant knowledge for the psychological therapies. *Clinical Psychology & Psychotherapy. 10*, 319–327

Barrett-Lennard, G. T. (1962). Dimensions of therapist response as a causal factor in therapeutic change. *Psychological monographs: General and Applied*, *76,* 43 (Whole No. 562).

Barrett-Lennard, G. T. (1964). *The relationship inventory forms: OS-M-64, OS-F-64 and MO-M-64 plus MO-F-64.* Unpublished manuscript, University of New England Australia.

Beck, A.T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty Years of evaluation. *Clinical Psychology Review, 8,* 77-100.

Benjamin, L. S. (1996). Introduction to the special section on structural analysis of social behavior. *Journal of Consulting and Clinical Psychology, 64*(6) 1203-1212.

Benjamin, L. S. (2005). Interpersonal theory of personality disorders: The structural analysis of social behavior and interpersonal reconstructive therapy. In M. F. Lenzenweger & J. F. Clarkin (Eds.), *Major Theories of Personality Disorder* (2nd ed., pp. 157-230). New York: Guilford Press

Benson, J. (1998). Developing a strong program of construct validation: a test anxiety example. *Educational Measurement: Issues and Practice. 17*(1), 10-17

Berzon, B., Pious, C. & Farson, R. E. (1963). The therapeutic event in group psychotherapy; study of subjective reports by group members. *Journal of Individual Psychology, 19*, 201-12

Bibring, E. (1937). Contribution to the symposium on the theory of the therapeutic results of psycho-analysis. *International Journal of Psycho-Analysis*, *18*, 170-189.

Bloch, S., & Reibstein, J. (1980). Perception by patients and therapist of therapeutic factors in group psychotherapy. *British Journal of Psychiatry*, *137*, 274–278.

Bloch, S., Reibstein, J., Crouch, E., Holroyd, P. & Themen, J. (1979). A method for the study of therapeutic factors in group psychotherapy. *British Journal of Psychiatry, 134,* 257–263.

Bohart, A.C., Elliott, R., Greenberg, L.S. & Watson, J.C. (2002). Empathy. In J. Norcross (Ed.), *Psychotherapy Relationships That Work* (pp. 89-108). New York: Oxford University Press.

Bond, T. & Fox, C. (2007). *Applying the Rasch Model: Fundamental measurement in the Human Sciences* (2nd ed.). New Jersey: Lawrence Erlbaum Associates.

Bordin, E.S (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy, 16,* 252-260.

Bordin, E.S. (1994). Theory and research on the therapeutic working alliance: New Directions. In A. O. Horvath and L. S. Greenberg (Eds.), *The Working Alliance. Theory, Research, and Practice* (pp. 13-37). New York: Wiley.

Bourdeau, M. (2011). Auguste Comte. In E. N Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Summer 2011 Edition)*. Retrieved on 17th Feb 2012 from http://plato.stanford.edu/archives/sum2011/entries/comte/

British Association for Counselling and Psychotherapy. (2011). *Research investigating the effectiveness of counselling and psychotherapy: The importance of research in counselling and psychotherapy*. Retrieved on June 1st 2011 from http://www.bacp.co.uk/research/counselling_effectiveness.php#Glover_2010

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 3*, 296–322.

Buber, M. (1958). *I and thou* (R.G. Smith, Trans.). New York: Charles Scribner's Sons.

Budgell, R. (1995). *Being touched through space*. Unpublished dissertation of School of Psychotherapy and Counselling, Regents College, London.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validity in the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105

Campos, J. A. D. B., Zucoloto, L. M., Bonafe, F. S. S., Jordani, P. C, & Maroco, J. (2011). Reliability and validity of self-reported burnout in college students: A cross randomized comparison of paper-and-pencil vs. online administration. *Computers in Human Behavior, 27,* 5.

Cardinal, D., Hayward, J. & and Jones, G. (2004). *Epistemology: The theory of knowledge.* London: Hodder Education

Castonguay, L. G., Boswell, J. F., Zack, S. E., Baker, S., Boutselis, M. A., Chiswick, N. R., Damer, D. D., Hemmelstein, N. A., Jackson, J. S., Morford, M., Ragusea, S. A., Roper, J. G., Spayd, C., Weiszer, T., Borkovec, T. D. & Holtforth, M. G. (2010). Helpful and hindering events in psychotherapy: A practice research network study. *Psychotherapy: Theory, Research, Practice, Training*, *47(*3), 327-344

Connelly, A. E. (2009). *Trainee psychotherapists and relational depth.* Unpublished MA Dissertation. The Sherwood Institute Nottingham and University of Birmingham.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd ed.). New Jersey: Lawrence Erlbaum.

Cook, D.A. & Beckham, T.J. (2006). Current concepts in validity and reliability for psychometric instruments: Theory and application. *The American Journal of Medicine*, *119*, 2

Coolican, H. (1999). *Aspects of psychology: Research methods and statistics.* London: Hodder and Stoughton

Cooper, M. (2005). Therapists' experiences of relational depth: A qualitative interview study. *Counselling and Psychotherapy Research, 5*(2), 87-95

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

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281-302.

Cronbach, L.J., Rajaratnman, N. & Gleser, G.C. (1963). Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology, 16*, 137–163.

Cronbach, L. J. (1988). Five perspectives on validity argument. In W. Wainer & H. Braun (Eds.), *Test validity* (pp. 3-17). New Jersey: Lawrence Erlbaum.

De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: The Guildford Press.

Denniston, M. M., Brener, N. D., Kann, L., Eaton, D. K., McManus, T., Kyle, T. M., Roberts, A. M., Flint, K. H., & Ross, J. G. (2010). Comparison of paper-and-pencil versus web administration of the Youth Risk Behavior Survey (YRBS): Participation, data quality, and perceived privacy and anonymity. *Computers in Human Behaviour, 26,* 1054-160.

Denscombe, M. (2006). Web-based questionnaires and the mode effect - An evaluation based on completion rates and data contents of near-identical questionnaires delivered in dfferent modes. *Social Science Computer Review, 24*, 246-254.

Derogatis, L.R., Rickels, K., & Rock, A.F. (1976). The SCL-90 and the MMPI: A step in the validation of a new self-report scale. *British Journal of Psychiatry, 128*, 280–289.

De Vellis, R.F. (2003). *Scale development: theory and applications*. United Kingdom: Sage Publications Ltd.

Elliott, R. (1985). Helpful and nonhelpful events in brief counseling interviews: An empirical taxonomy. *Journal of Counseling Psychology, 32*(3), 307-322.

Elliott, R., Bohart, A.C., Watson, J. C. & Greenberg, L. S. (2011). Empathy. In Norcross, J. (Ed.), *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd ed., pp. 132-152). New York: Oxford University Press.

Elliott, R., Fox, C. M., Beltyukova, S. A., Stone, G. E., Gunderson, J. & Zhang, X. (2006). Deconstructing therapy outcome measuremment with Rasch analysis of a measure of general clinical distress: The Symptom Checklist-90-Revised. *Psychological Assessment, 18*(4), 359-372.

Elliott, R., Shapiro, D.A., Firth-Cozens, J., Stiles, W.B., Hardy, G., Llewelyn, S.P, & Margison, F. (1994). Comprehensive process analysis of insight events in cognitive-behavioral and psychodynamic-interpersonal therapies. *Journal of Counseling Psychology, 41*, 449-463.

Elliott, R., Shapiro, D. A., & Mack, C. (1999). Simplified personal questionnaire procedure manual. Toledo, Ohio: University of Toledo, Department of Psychology.

Elliott R., Watson J. C., Goldman, R. N., & Greenberg L., S. (2004). *Learning emotion-focused therapy: The process-experiential approach to change.* Washington, DC: American Psychological Association.

Eugster, S. L., & Wampold, B.E. (1996). Systematic effects of participant role on evaluation of the psychotherapy session. *Journal of Consulting and Clinical Psychology*, *64,* 1020-1028.

Evans, C., Connell, J., Barkham, M., Margison, F., Mellor-Clark, J., McGrath, G. & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry*, *180*, 51–60.

Ewing M. T., Salzberger T., & Sinkovics R. R. (2009). Confirmatory factor analysis vs. Rasch approaches: Differences and measurement implications. *Rasch Measurement Transactions*. Retrieved on 16th Dec 2011 from http://www.rasch.org/rmt/rmt231g.htm

Farber, B. A. & Doolin, E. M. (2011). Positive Regard and affirmation. In Norcross, J. (Ed.), *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd ed., pp. 168-186). New York: Oxford University Press.

Fenton, L., Cecero, J. J., Nich, C., Frankforter, T. L. & Carroll, K. M. (2001). Perspective is everything: The predictive validity of six working alliance instruments. *Journal of Psychotherapy Practice Research, 10*(4), 262-268.

Freire, E. & Cooper, M. (2007). The Strathclyde Inventory: validation of a person-centred outcome measure. Paper presented at BACP Research Conference, Birmingham, UK.

Freud, S. (1958a). The dynamics of transference. In J. Strachey (Ed. & Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 12, pp. 99—108). London: Hogarth Press. (Original work published 1912))

Freud, S. (1958b). On the beginning of treatment: Further recommendations on the technique of psychoanalysis. In J. Strachey (Ed. & Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 12, pp. 122—144). London: Hogarth Press. (Original work published 1913).

Frey, B. B., Petersen, S., Edwards, L. M., Pedrotti, J. T., & Peyton, V. (2005). Item-writing rules: Collective wisdom. *Teaching and Teacher Education, 21,* 357-364.

Gaston, L., & Marmar, C. (1994). The California Psychotherapy Alliance Scales. In A. O. Horvath & L. S. Greenberg (Eds.), *The working alliance: Theory, research and practice*. New York: Wiley.

Gelso, C. J. (2011). *The real relationship in psychotherapy.* Washington: American Psychological Association

Gelso, C. J. & Carter, J. A. (1994). Components of the psychotherapy relationship: Their interaction and unfolding during treatment. *Journal of Counseling Psychology, 41*, 296-306.

Glymour, C. & Eberhardt, F. (2011). Hans Reichenbach. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Summer 2011 Edition)*. Retrieved on 20th Dec 2011 from http://plato.stanford.edu/archives/sum2011/entries/reichenbach/

Goodwin, L. D. & Leech, N.L. (2003). The meaning of validity in the New Standards for Educational and Psychological Testing: Implications for measurement courses. *Measurement and Evaluation in Counseling and Development, 36,* 181-191

Greenson, R. (1965). The working alliance and the transference neurosis. *Psychoanalytic Quarterly*, *34*, 155-181.

Gregory, R. J. (2007). *Psychological testing: History, principles, and applications* (5th ed.). Boston: Allyn & Bacon.

Kolden, G. K., Klein, M. H., Wang, C-C & Austin, S. B. (2011). Congruence/Genuineness. In Norcross, J C. (Ed.), *Psychotherapy relationships that work* (2nd ed., pp. 187-202). New York: Oxford University Press.

Gullikson, H. (1950). *Theory of Mental Tests.* New York: Wiley

Gurman, A. S. (1977). The patient's perception of the therapeutic relationship. In A. S. Gurman & A. M. Razin (Eds.), *Effective sychotherapy: A handbook of research* (pp. 503-543). New York: Pergamon.

Hanna, R.C., Weinberg, B., Dant, R. P. & Berger (2005). Do internet-based surveys increase personal self-disclosure? *Database marketing & Customer Strategy Management, 12*(4), 342-356.

Hart, T. (1997). Transcendental Empathy is the therapeutic encounter. *The Humanistic Psychologist, 25*(3), 245-270.

Hartley, D., & Strupp, H. (1983). The therapeutic alliance: Its relationship to outcome in brief psychotherapy. In J. Masling (Ed.), *Empirical studies of psychoanalytic theories* (pp. 1-27). Hillsdale, NJ: Erlbaum.

Hatcher, R. L., & Gillaspy, J. A. (2006). Development and validation of a revised short version of the Working Alliance Inventory. *Psychotherapy Research, 16*(1), 12-25

Herrero, J. & Meneses, J. (2006). Short Web-based versions of the perceived stress (PSS) and Center for Epidemiological Studies-Depression (CESD) Scales: a comparison to pencil and paper responses among Internet users. *Computers in Human Behaviour, 22*, 830-846.

Hewson, C. (2003). Conducting research on the internet. *The Psychologist 16*(6), 290-293

Horvath, A. O. (1994). Empirical validation of Bordin's pantheoretical model of the alliance: The Working Alliance Inventory perspective. In Horvath, A. O. & Greenberg, L. S. (Eds.), *The working alliance: Theory, research, and practice* (pp. 109-128). New York: Wiley.

Horvath, A. O. & Bedi, R.P. (2002). The alliance. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp. 37-69) Oxford: Oxford University Press.

Horvath, A. O., Del Re, A., Fluckiger, C., & Symonds, D. B. (2011). Alliance in individual psychotherapy. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd. ed., pp. 25-69). New York: Oxford University Press.

Horvath, A. O., Del Re, A., Fluckiger, C., & Symonds, D. B. (2011). Alliance in Individual Psychotherapy. *Psychotherapy*, *48*(1), 9-16

Horvath, A. O. & Greenberg, L. S. (1986). Development of the Working Alliance Inventory. In L. S. Greenberg & W. M. Pinsoff (Eds.), *The psychotherapeutic process: A research handbook* (pp. 529-556). New York: Guilford.

Horvath, A. O. & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology, 36*, 223-233.

Horvath, A. O. & Luborsky, L. (1993). The role of the therapeutic alliance in psychotherapy. *Journal of Consulting and Clinical Psychology, 61,* 561–573.

Horvath, A. O. and Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy. *Journal of Counseling Psychology, 38*, 139-149.

Howitt, D. & Cramer, D. (2008). *Introduction to statistics in psychology* (4th ed.). Essex: Pearson Education

Joinson, A. J. (1999). Social desirability, anonimity, and Internet-based questionnaires. *Behavior Research Methods, Instruments, & Computer, 31*(3), 433-438.

Joinson, A. N., Woodley, A. & Reips, U. (2007). Personalization, authentication and self-disclosure in self-administered internet surveys. *Computers in Human Behaviour*, *23*, 275-285.

Kaplan, R. M. & Saccuzzo, D. P. (2009). *Psychological Testing: Principles Applications and Issues.* Belmont, CA: Wadsworth.

Karver, M. S., Handelsman, J. B., Fields, S., & Bickman, L. (2006). Meta-anlysis of therapeutic relationship variables in youth and family therapy: The evidence for different relationship variables in the child and adolescent treatment outcome literature. *Clinical Psychology Review*, *26*, 50-65

Keys, S. (2007). *Love in therapy*. Mary Kilborn Lecture. Retrieved August 14th 2011, from http://www.strath.ac.uk/counsunit/features/presentations

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

Knox, R. (2007). *Clients experiences of relational depth in person-centred counselling.* Unpublished manuscript.

Knox, R. (2008). Clients' experience of relational depth in person-centred counselling. *Counselling and Psychotherapy Research, 8*(3), 182-189.

Knox, R. (2011). *Clients' experiences of relational depth.* Unpublished PhD dissertation, University of Strathclyde, Glasgow.

Knox, R. & Cooper, M. (2010). Relationship qualities that are associated with moments of relational depth: The client's perspective. *Person-Centered and Experiential Psychotherapies, 9*(3), 236-256.

Kuder, G. F. & Richardson, M. W. (1937). The Theory of the Estimation of Test Reliability. *Psychometrika, 2*, 151-160.

Kvale, S. (1995). The Social Construction of Validity. *Qualitative Inquiry 1*(1), 19-40.

LaCrosse, M. B. (1980). Perceived counselor social influence and counseling outcomes: Validity of the Counselor Rating Form. *Journal of Counseling Psychology*, *27*, 320-327.

Lambert, M., J. & Barley, D., E. (2001). Research Summary on the therapeutic relationship and psychotherapy outcome. *Psychotherapy*, *38*(4), 357-361.

Leung, J. (2009). A quantitative online study exploring the factors associated with the experience and perceived importance of relational depth. Unpublished manuscript.

Linacre, J. M. (1997). KR-20 or Rasch reliability: Which tells the "truth"? *Rasch Measurement Transactions, 11*(3), 580-581. Retrieved 30th May, 2011 from http://www.rasch.org/rmt/contents.htm

Linacre, J. M. (1998). Structure in Rasch residuals: Why principal components analysis (PCA)? *Rasch Measurement Transactions, 12*(2), 636. Retrieved 30th May, 2011 from http://www.rasch.org/rmt/contents.htm

Linacre, J. M. (2011). A user's guide to winsteps ministep Rasch-model computer programs (Version 3.71). Retrieved 27th March, 2011, from http://www.rasch.org/software.htm

Llewelyn, S. (1988). Psychological therapy as viewed by clients and therapists. *British Journal of Clinical Psychology, 27*, 223-238.

Locke, J. (1976). An essay concerning human understanding. Woozley, A. D. (Ed.). Glasgow: Collins

Luborsky, L. (1976). Helping alliances in psychotherapy. In J. L. Cleghhorn (Ed.), *Successful psychotherapy* (pp. 92–116). New York: Brunner/Mazel.

Luborsky, L. (1994). Therapeutic alliances as predictors of psychotherapy outcomes: Factors explaining the predictive success. In A. O. Horvath & L. S. Greenberg (Eds.), *The Working Alliance: Theory, Research and Practice*. New York: Wiley

Luborsky, L., Barber, J.P., Siqueland, L., Johnson, S., Najavits, L.M., Arlene, F. & Daley, D. (1996). The Revised Helping Alliance Questionnaire (Haq-II). *Journal of Psychotherapy Practice and Research, 5,* 260-271.

Luborsky, L., Crits-Christoph, P., Alexander, L., Margolis, M. & Cohen, M. (1983). Two helping alliance methods for predicting outcomes of psychotherapy: A counting ssigns vs a global rating method. *The Journal of Nervous and Mental Disease, 171* (8), 480-491.

Lyne, K. D., Barrett, P., Evans, C. and Barkham, M. (2006). Dimensions of variation on the CORE-OM. *British Journal of Clinical Psychology 45*(2), 185-203

Margison, F., Barkham, M., Evans, C., McGrath, G., Mellor-Clark, J., Audin, K., & Connell, J. (2000). Measurement and psychotherapy: evidence-based practice and practice-based evidence. *British Journal of Psychiatry*, *177*, 123–130.

Markie, P. (2012). Rationalism vs. Empiricism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Spring 2012 Edition)*. Retrieved on March 1st 2012 from http://plato.stanford.edu/archives/spr2012/entries/rationalism-empiricism/

Marmar, C. R., Weiss, D. S., & Gaston, L. (1989). Towards validation of the California Therapeutic Alliance Rating System. *Journal of Consulting and Clinical Psychology*, *1*, 46-52.

Martin, R. J. (2010). *Epistemology*. Oxford: Oneworld Publications

Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 68*, 438-450.

Marsh, H. W. (1993). The multidimensional structure of physical fitness: Invariance over gender and age. Research Quarterly for Exercise and Sport, *64*(3), 256-273.

Marziali, E. (1984). Three viewpoints on the therapeutic alliance scales: Similarities, differences, and associations with psychotherapy outcome. *Journal of Nervous and Mental Disease, 172,* 417-423.

McMillan, M. & McLeod, J. (2006). Letting go: The client's experience of relational depth. *Person-Centered and Experiential Psychotherapies, 5,* 277-292.

Mearns, D. (1996). Working at relational depth with clients in person-centred therapy. *Counselling, 7*, 307-311.

Mearns, D. (1997). *Person-centred counselling training*. London : Sage.

Mearns, D. (2003). *Developing person-centred counselling* (2nd edn.). London: Sage

Mearns, D. (2009). *Relationship needed but also feared: The challenges of relational depth.* Retrieved on May 19th 2011 from www.davemearns.com/**RELATIONSHIP**.pdf

Mearns, D., & Cooper, M. (2005). *Working at relational depth in counselling and psychotherapy*. London: Sage

Mearns, D. & Thorne, B. (2000). *Person-centred therapy today: New frontiers in therapy and practice.* London: Sage

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist, 35*, (11), 1012-1027

Messick, S. (1990). *Validity of test interpretation and use*. Princeton, NJ: Educational Testing Service.

Messick, S. (1995). Validity of psychological assessment: Validity of inferences from persons' responses and performances as scientific enquiry into score meaning. *American Psychologist, 50*(9), 741-749.

Moodley, R. & Lijtmaer, R. (2007). Beyond the boundary of analytic revelation: Culture, self-disclosure (analyst and patient) and anonimity. *Psychologist – Psychoanalyst, 27,* 49-51. American Psychological Association.

Moreno, J. K. Fuhriman, A., & Hileman, E. (1995). Significant events in a psychodynamic psychotherapy group for eating disorders. *Group, 19,* 56-62.

Morgan, R. W., Luborsky, J., Crits-Christoph, P., Curtis, H. & Solomon, J. (1982). Predicting the outcomes of psychotherapy using the Penn Helping Alliance rating method. *Archives of General Psychiatry, 39*, 397-402.

Murphy, D. (2010). Psychotherapy as a mutual encounter: A study of therapeutic conditions. Unpublished manuscript.

Murzi, M. (2001) (Updated 2005). Hans Reichenbach (1891-1953). *Internet Encyclopedia of Philosophy*. Retrieved on 20 Dec 2011 from http://www.iep.utm.edu/reichenb/

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology. 3*(1), 1-18.

Ogden, T. (1994a). *Subjects of analysis*. London: Karnac Books.

Ogden, T. (1994b). The analytic third: Working with intersubjective clinical facets. *The International Journal of Psycho-Analysis, 75*, 3-19

Ogden, T. (2004). The Analytic Third: Implications for Psychoanalytic Theory and Technique. *The International Journal of Psycho-Analysis, 73*, 167-195

Olsen, L. W. (2003). Essays on George Rasch and his Contributions to Statistics. PhD Thesis, Institute Of Economics, University of Copenhagen Retrieved on 5th Jan 2011 from http://www.rasch.org/olsen.htm

O'Malley, S. S., Suh, C. S. & Strupp, H. H. (1983). *Journal of Consulting and Clinical Psychology*, *51*, 581-586.

Orlinsky, D.E., & Howard, K. I. (1977). The therapist's experience of psychotherapy. In A. S. Gurman & A. M. Razin (Eds.), *Effective psychotherapy: A Handbook of research.* (pp. 566-589). New York: Pergamon.

Orlinsky, D.E., and Howard, K. I. (1986). The psychological interior of psychotherapy: Explorations with the Therapy Session Report Questionnaires. In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook*. (pp. 477-501). New York: Guilford.

Orlinsky, D.E., & Rønnestad, M.H. (2005). *How psychotherapists develop: A study of therapeutic work and professional growth*. Washington, DC: American Psychological Association.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp.17-59). New York: Academic Press.

Pembroke, N. (2006). *Renewing pastoral practice: Trinitarian perspectives on pastoral care and counselling.* Aldershot: Ashgate

Pearson, K. (1901). *Mathematical contributions to the theory of evolution*. London: Dulau & Co.

Prendergast, J. J., Fenner, P. & Krystal, S. (2003). *The sacred mirror: Nondual wisdom and psychotherapy.* Minnesota: Paragon House

Raat, H., Mangunkusumo, R. T., Landgraf, J. M., Kloek, G. & Brug, J. (2007) Feasibility, reliability, and validity of adolescent health status measurement by the Child Health Questionnaire Child Form (CHQ-CF): internet administration compared with the standard paper version. *Quality of Life Research, 16*, 675-685

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, *IV* (pp. 321–334). Berkeley: University of Chicago Press.

Rennie, D.L., Phillips, J.R., Quartaro, G.K. (1988). Grounded theory: A promising approach to conceptualization in psychology? *Canadian Psychology, 29*, 139-150

Risko, E. F., Quilty, L. C. & Oakman, J. M. (2006). Socially Desiriable Responding on the Web: Investigating the Candor Hypothesis. *Journal of Personality Assessment, 87*(3), 269-276

Rogers, C. R. (1951). *Client-Centered Therapy.* Constable: London

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21*, 95-103.

Rogers, C. R. (1959). A theory of therapy, personality and interpersonal relationships, as developed in the Client-Centred framework. In S. Koch (Ed.), *Psychology: A Study of Science* (pp. 184-256). New York: McGraw-Hill

Rogers, C. R. (1961). *On becoming a person*. Constable: London

Rogers, C. R. (1980). *A way of being*. Boston: Houghton Mifflin

Rowan, J. (1998). Linking: Its place in therapy. *International Journal of Psychotherpy, 3*(3), 245-54.

Rowan, J. (2005). *The transpersonal: Spirituality in psychotherapy and counselling* (2nd Ed.). London: Routledge

Rowan, J. (2010). *Personification: Using the dialogical self in psychotherapy and counselling*. East Sussex: Routledge

Rowe, D. & Mahar, M. T. (2006). Validity. In T. Wood & W. Zhu (Eds.). *Measurement theory and practice in kinesiology* (pp. 9-26). Champaign, IL: Human Kinetics.

Ryan, A. B. (2006). Post-positivist approaches to research. In M. Antonesa, H. Fallon, A. B. Ryan, A. Ryan, & T. Walsh, *Researching and writing your thesis: A Guide for Postgraduate Students* (pp. 12-26). Mace: Maynooth Adult and Community Education

Salvio M., Beutler L.E., Wood J.M., & Engle, D. (1992). The strength of the therapeutic alliance in three treatments for depression. *Psychotherapy Research 2*(1), 31–36.

Schmidt, W.C. (1997). World Wide Web survey research: Benefits, potential problems, and solutions. *Behaviour Research Methods, Instruments and Computers, 29*, 270-273.

Sick, J. R. (2008a). Rasch measurement in language education Part 1. *Shiken, 12*(1), 1-6. Retrieved 20th April, 2011 from http://jalt.org/test/sic_1.htm

Sick, J. R. (2008b). Rasch measurement in language education Part 2: Measurement Scales and Invariance. *Shiken, 12*(2), 26-31. Retrieved 20th April, 2011 from http://jalt.org/test/sic_2.htm

Sick, J. R. (2011). Rasch measurement in language education: Part 6: Rasch Measurement and Factor Analysis. *Shiken, 15*(1), 15-17. Retrieved 20th Dec, 2011 from http://jalt.org/test/PDF/Sick6.pdf

Simmons J, Roberge L, Kendrick, B. S. (1995). The interpersonal relationship in clinical practice: the Barrett-Lennard Relationship Inventory as an assessment instrument. *Evaluation and the Health Professions*, *18*, 103–112.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*(1), 72-101

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 3*, 271–295.

Sterba, R. (1934). The fate of the ego in analytic therapy. *International Journal of Psychoanalysis, 15*, 117-126.

Stern, D. N. (2004). *The present moment in psychotherapy and everyday life*. New York : W.W. Norton.

Steup, M. (2011). Epistemology. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Winter 2011 Edition)*. http://plato.stanford.edu/archives/win2011/entries/epistemology/Retrieved 20th January 2012.

Stevens, S.S. (1946). On the theory of scales of measurement. *Science*, 103, 677-680.

Stevens, S.S. (1968). Measurement, statistics and the schemapiric view: Like the faces of Janus, science looks two ways-towards schematics and empirics. *Science, 161*, 849-856.

Stiles, W.B., Gordon, L. E., & Lani, J. A. (2002). Session evaluation and the Session Evaluation Questionnaire. In G. S. Tryon (Ed.), *Counseling based on process research: Applying what we know* (pp. 325-343). Boston, MA: Allyn & Bacon

Stiles, W. B. & Snow, J. S. (1984). Counseling session impact as viewed by novice counselors and their clients. *Journal of Consulting and Clinical Psychology, 31*, 3-12.

Stiles, W. B., Shapiro, D., & Elliott, R. (1986). Are all psychotherapies equivalent? *American Psychologist*, *32*, 752-760.

Stone, L. (1961). *The psychoanalytic situation: An examination of its development and essential nature.* Oxford, England: International University Press.

Suh, C. S., Strupp, H. H. & O'Malley, S. S. (1986). The Vanderbilt process measures: The Psychotherapy Process Scale (VPPS) and the Negative Indicators Scale (VNIS). (In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook* (pp. 285—323). New York: Guilford Press)

Surveymonkey. (2006). Surveymonkey. Retrieved July 2006 http://www.surveymonkey.com (version used was the one before May 26, 2010 and not the current version (at time of writing) which is from this date).

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th Ed.). Boston: Allyn and Bacon.

Tennant, A.& Pallant, J.F. (2006). Unidimensionality matters! (A tale of two Smiths?) *Rasch Measurement Transactions, 20(1)*, 1048-51.

Thomas, R., Stephenson, S. & Loewenthal., D. (2006). UKCP *Report on practice-based evidence (PBE)*. London: UKCP.

Thorndike, E. L. (1916). *An introduction to the theory of mental and social measurements* (2nd ed.). New York: Teachers College Press.

Thorne, B. (1991). Person-centred counselling: *Therapeutic and spiritual dimensions.* London: Whurr.

Tracey, T. J., & Kokotovic, A. M. (1989). Factor structure of the Working Alliance Inventory. *Psychological Assessment, 1*, 207-210

Trochim, W. M. K., (2001). *The research methods knowledge base* (2nd ed.). Cincinnati, OH: Atomic Dog. Publishing.

Timulak, L. (2007). Identifying core categories of client-identified impact of helpful events in psychotherapy: A qualitative meta-analysis. *Psychotherapy Research 17*(3), 305-314.

Timulak, L. (2010). Significant events in psychotherapy: An update of research findings. *Psychology and Psychotherapy: Theory, Research and Practice*, *83*, 421-447.

Timulak, L., & Elliott, R. (2003). Empowerment events in process-experiential psychotherapy of depression. *Psychotherapy Research, 13*, 443–460.

Tracey, T. J., & Kokotovic, A. M. (1989). Factor structure of the Working Alliance Inventory. *Psychological Assessment*, *1*, 207-210.

Uebersax, J. S. (2000). Latent Trait Analysis and Item Response Theory (IRT) Models. *Latent Structure Analysis.* Retrieved on 26th February 2012 from http://www.john-uebersax.com/stat/index.htm.

Vallejo, M. A., Mananes, G., Comeche, M. I. & Diaz, M I. (2008). Comparison between administration via Internet and paper-and-pencil administration of two clinical instruments: SCL-90-R and GHQ-28. *Journal of Behavior Therapy and Experimental Psychiatry*, *39*, 201-208

Van de Looij-Jansen, P. M. & De Wilde, E. J. (2008). Comparison of Web-based versus paper-and-pencil self-administered questionnaire: Effects on health indicators in dutch adolescents. *Health Research and Educational Trust, 43,* 1708-1721

Wagner, J, & Elliott, R. (2001). *The Simplified Personal Questionnaire*. Manuscript submitted for publication, Department of Psychology, University of Toledo.

Watson, J. & Geller, S. M. (2005). The relation among the relationship conditions, working alliance, and outcome in both process-experiential and cognitive –behavioral psychotherapy. *Psychotherapy Research, 15*(1-2), 25-33.

Wiggins, S. (2007). Developing an Inventory Designed to Assess Relational Depth. Unpublished master's thesis, University of Strathclyde, Glasgow, UK.

Wiggins, S., Elliott, R. & Cooper, M. (2012). The prevalence and characteristics of relational depth events in psychotherapy. *Psychotherapy Research, 22*(2), 139-158

Wilber, K. (1980). *The Atman project*, Wheaton, U.S.A: Theosophical Publishing House.

Wilders, S. (2007). Relational depth and the person-centred approach. *Person-Centred Quarterly,* February, 1- 4.

Wright, B.D. (1997). Measurement for social science and education: History of social science measurement. *Institute of Objective Measurement: Research Memoranda.* Retrieved 30th May, 2011 from http://www.rasch.org/memo62.htm

Wright, B.D. (1999). Common sense for measurement. *Rasch Measurement Transactions. 13*(3), 704.

Wright, B. D. & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press

Wright, K.B. (2005). Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication, 10*(3), article 11. Retrieved May 18th 2011 from http://jcmc.indiana.edu/vol10/issue3/wright.html

Yan, Z. (2009). *The Development of a Rasch Measurement Physical Fitness Scale for Hong Kong Primary School-Aged Students*. PhD thesis, James Cook University, Australia. Retrieved 19th February 2011 from http://eprints.jcu.edu.au/5569/

Yun, J. & Ulrich, D.A. (2002). Estimating measurement validity: A tutorial. *Adapted Physical Activity Quarterly, 19*, 32-47

Zetzel, E. R. (1956). Current concepts of transference. *International Journal of Psycho-Analysis, 37*, 369-375.

Zetzel, E.R. (1965). The theory of therapy in relation to a developmental model of the psychic apparatus. *International Journal of Psycho-Analysis*, *46*, 39.

Zuroff, D.C. and Blatt, S.J. (2006). The therapeutic relationship in the brief treatment of depression: Contributions to clinical improvement and enhanced adaptive capacities. *Journal of Consulting and Clinical Psychology*, *74*, 130-140

Appendices

Appendix 1: Relational Depth Inventory

Important event or experience during counselling

On this page you are asked about a particular important event or experience which you might have encountered during a therapy session.

Please take a minute to think back over your relationship so far with this therapist or client. Of the events which have occurred so far, select one that stands out in your mind as particularly important. This event or experience can be positive, negative or neither, as long as it felt important to you. Please briefly describe this event below in not more than about 3 sentences.
If you are a counsellor, please do not include details that you feel might compromise your client's confidentiality; instead you might concentrate on a brief description of the process and how you felt at that moment.

Now, with this particular event in mind, please rate how accurately each of the items below fits your experience of this event:

During this important event, I felt…

Please tick in the appropriate box to indicate your answer.	not at all	slightly	some-what	very much	completely
1. liberated					
2. we had shared things in common					
3. a sense of flow (a sense of smooth and continuous change in myself)					
4. there was give and take					
5. intense feelings					
6. the other person respected me					
7. an atmosphere that was meaningful					
8. a sense of being in the moment					
9. the other person valued me					
10. spiritual					
11. love					
12. a meeting of minds					
13. opposing feelings at the same time					
14. a sense of privacy in the relationship					
15. centred on the present					
16. energised					
17. courageous					
18. trust (in the relationship)					
19. safe					
20. in an altered state					
21. revitalised					
22. scared					
23. magical					
24. the both of us were connected in some					

way					
25. weird feelings					
26. angry					
27. Paradoxical (seemingly contradictory but nonetheless possibly true)					
28. The other person trusted me					
29. an inexplicable atmosphere					
30. a silent atmosphere					
31. self-value					
32. sexual					
33. an awesome atmosphere					
34. vulnerable					
35. empathy for the other person					
36. expansive (unrestrained)					
37. the other was empathic towards me					
38. happy					
39. a still atmosphere					
40. I was transcendent (going beyond my ordinary limits)					
41. mutuality					
42. the other person was real					
43. I was at one with the other					
44. I was there for the other					
45. immersed					
46. intimacy					
47. in touch with self					
48. spontaneous					
49. value of other					
50. equality					
51. a timeless atmosphere					
52. respect for other					
53. open					
54. I was being real					
55. a unique atmosphere					
56. exhilarated					
57. trust of other					
58. excited					
59. empowered					
60. a transformative atmosphere					
61. other was available to me					
62. soulful (deep feelings and/or emotions)					
63. aware of experience					
64. enlightened					

Appendix 2: Relational Depth Event Content Rating Scale

What is a Relational Depth Event?

A relational depth event might be described as a moment, a series of moments, an experience or feeling where the participant appears to be describing ***'a state of profound contact and engagement between two people [between themselves and the other person], in which each person is fully real with the Other, and able to understand and value the Other's experiences at a high level'*** (Mearns & Cooper, 2005).

Instructions

Assign a score from 0 to 3 to the narrative descriptions of events, depending how clearly relational depth is deemed present (0: "clearly not present"; 1: "probably not present"; 2: "probably present or present but not strongly"; 3: "clearly or strongly present").

If, after rating, there are discrepancies between raters (e.g. if one rater assigned a score of 1 and another a score of 2), the final score should be an average (e.g. in this case to 1.5). If discrepancies are more than 1 (e.g. one rater assigned a 1 and the other a 3), raters should discuss the case until consensus is reached.

Scoring: Version 1:

3 – Relational Depth event clearly or strongly present. This would typically be an event which clearly suggested a 'two-wayness' and high levels of depth.

2 – Relational Depth probably present (not certain or not strongly present; implied or inferred or strongly suspected even though not explicit in the description). Evidence of two-wayness & depth but not certain, explicit, or clear.

1 -- Probably not relational depth (some evidence, but not enough; a "hint" of the variable). There might be evidence of sharing and emotional language but no evidence of connectedness or depth.

0 -- Clearly not relational depth. No evidence at all of two-wayness, emotional language or sharing.

Example descriptions and scoring

Score/rating of 3

'I was deeply understood as I have never been before. It was the first time that someone really understood me. It was magical, powerful and liberating..' (Rate 3)

'My therapist acknowledged the full extent of the issues I have been facing. She intuitively extended her hands to me and I hugged her. At this point in the therapy we became closer and a level of understanding was thereafter assumed'. (Rate 3)

'A relationship within which a 'pause' of silence there was a powerful sense of great depth and feeling shared in the physical space between us'. (Rate 3)

Score/rating of 2

'The therapist challenged me on the way I had been using 'merging' to avoid facing an issue. Although I didn't like it, I immediately felt the trust he had put in me in

taking the risk to make the challenge. We both shared our feelings in the present and agreed a contract for the future'. (Rate 2)
'The client has expressed great distress and lack of hope, as well as fear. I feel very touched by her despair and state a real longing to be of help to her but now really knowing what to offer, other than my conviction that she is a person of worth...'. (Rate 2)
'me getting pregnant opened new topics and made me feel more close to my therapist who is a mother and also shared personal and professional topics with me…'. (Rate 2)
A session where I was able to be myself as a small child and to gain the ability to comfort myself without feeling embarrassed by the process and where I felt emotionally held by the therapist…'. (Rate 2)

Ratings of 1

Gaining deeper insight into the way I experience stress by sharing and exploring this with my therapist (Rate 1)
Experienced clarity and contact with a clear definition, connectedness with my own sense of spirituality which I had never experienced before (Rate 1)
'I was able, within the safety of the room, to allow myself to feel and fully experience the anger that I had been storing in my body for some time'. (Rate 1)

'Vision.. the client pains a completely new type of art after we talked about her changing using imagery in our words' (Rate 1)
'Client tells me that he is afraid of terminating therapy because that makes him losing a mentor, especially after losing his mom and dad'. (Rate 1)

Ratings of 0

'The moment when the client realised the reasons for her panic attacks were around her negative automatic thoughts'. (Rate 0)
'During some visualisations in the therapy, I had a strong sense of being in the wrong place in my life. At the same time, it felt like this was something I could handle, and tackle in the future, rather than now. This occurred towards the end of the period of therapy I undertook' (Rate 0)
'I had destructive, sadistic fantasies about the therapist. She invited me to explore this with her, but I did not dare to express my fantasies in detail'. (Rate 0).

Appendix 3: RDI-Revised

Below you are asked about a **particularly helpful moment or event** which you might have had during a therapy session. Please take a minute to think back over your relationship so far with this therapist. Of the events which have occurred so far, select a specific moment or event that stands out in your mind as particularly helpful. Please briefly describe this helpful moment or event below in a few sentences, and indicate about how long ago or in roughly what session it occurred.

Now, with this specific moment or event in mind, please rate how accurately each of the items below fits your experience. Please tick the appropriate box to indicate your answer.

During this specific moment or event…	not at all	slightly	some-what	very much	completely
1. I felt a sense of freedom	☐	☐	☐	☐	☐
2. There was give and take between my therapist and myself	☐	☐	☐	☐	☐
3. I felt my therapist respected me	☐	☐	☐	☐	☐
4. I felt I was 'living in the moment'	☐	☐	☐	☐	☐
5. I felt my therapist understood what it was like for me	☐	☐	☐	☐	☐
6. I felt in touch with my spiritual side	☐	☐	☐	☐	☐
7. My therapist and I both knew what was in each other's mind	☐	☐	☐	☐	☐
8. I felt more alive	☐	☐	☐	☐	☐
9. I felt a kind of magic happened	☐	☐	☐	☐	☐
10. I felt my therapist and I were both connected in some way	☐	☐	☐	☐	☐
11. I felt my therapist trusted me	☐	☐	☐	☐	☐
12. I felt my therapist was being genuine with me	☐	☐	☐	☐	☐
13. I felt the atmosphere was kind of awesome	☐	☐	☐	☐	☐
14. I felt I understood what it was like for my therapist	☐	☐	☐	☐	☐
15. I felt I was going beyond my ordinary limits	☐	☐	☐	☐	☐
16. My therapist and I felt close to each other	☐	☐	☐	☐	☐
17. I felt my therapist and I were equal	☐	☐	☐	☐	☐
18. I felt as if time had stopped	☐	☐	☐	☐	☐
19. I felt respect for my therapist	☐	☐	☐	☐	☐
20. I felt I was being genuine with my therapist	☐	☐	☐	☐	☐
21. I felt a sense of having my own power	☐	☐	☐	☐	☐
22. I felt my therapist was there for me	☐	☐	☐	☐	☐
23. I felt I had a better understanding of myself and/or others	☐	☐	☐	☐	☐
24. I felt a warm personal bond between myself and my therapist as fellow human beings	☐	☐	☐	☐	☐