

THE UNIVERSITY OF HULL

**ASSESSING PERSONALITY DISORDER
THROUGH STORYMAKING**

A reliability and validity study of the 6-Part Story Method

being a Thesis submitted
for the Degree of Doctor of Philosophy
in the University of Hull

by

Kim Dent-Brown, DipCOT, BSc (Hons), PG Dip (Dramatherapy)

April 2004

1.0 ABSTRACT

The 6-Part Story Method (6PSM) involves the participant creating and telling a fictional story which is then discussed with the interviewer. Stories were recorded from NHS mental health clinicians, and from adult mental health patients with and without a diagnosis of Borderline Personality Disorder (BPD). Statements describing the stories could be reliably scored by raters who were blind to the authorship of the stories. It was possible to use these statements to assemble a scale with good internal consistency, inter-rater reliability and test-retest reliability over a one-month period. The scale statements all related to the degree of negativity and failure expressed in the story. Ratings on this negativity scale given to stories from patients with a cluster B personality disorder diagnosis were significantly higher than the ratings given to stories from other participants.

The text of the story transcriptions was analysed using computerised text analysis programs, which detected some patterns of language use that appeared to distinguish reliably between stories from participants with BPD and other stories.

Participants were asked for their reactions to the 6PSM process, and their responses analysed using a Grounded Theory approach. This suggested that for most participants, the 6PSM works through increasingly close emotional identification with an initially distant and metaphorical main character.

2.0 ACKNOWLEDGEMENTS

I wish first and foremost to acknowledge the time and energy devoted to this study by the participants; the clinicians and patients who recorded stories, rated those of other people and undertook the lengthy diagnostic interviews needed. Many individuals gave up several hours of their own time for no reward other than a desire to help me and I thank them.

My supervisor, Professor Michael Wang, has been a great source of encouragement, support and friendly criticism since before this study started; as a PhD student this is the most important relationship there can be, and Mike has provided everything I could have wished for.

The former Northern and Yorkshire Regional Executive of the NHS provided funds via a Research Training Fellowship for the middle two years of the study; this enabled me to work full time on the project with no competing priorities, a rare and precious opportunity.

Before and after this period I was given generous support by the Research and Development department of my employing NHS Trust, Hull & East Riding Community Health and in particular its head, Tony Hostick. In the first year this gave me the platform to start the research and prepare the bid to Region for the Research Fellowship, and in the final year this has allowed me to finish writing the thesis and to plan the next stage of my research career.

In the very early stages I was encouraged to think about research into the 6PSM by Professor Glenys Parry of the University of Sheffield and Professor Rob Newell now of the University of Bradford. Once embarked on the study Professor Mooli Lahad and Dr Alida Gersie gave valuable encouragement and advice from their particular positions of expertise on the 6PSM.

The text analyses could not have been carried out without the help of Professor Philip Stone of Harvard University, who kindly provided a copy of his General Inquirer program. Professor James Pennebaker of the University of Texas carried out the analysis using his Linguistic Inquiry and Word Count program and offered much helpful advice and speculation.

My colleagues in the Psychotherapy Department of my NHS Trust have been very encouraging and tolerant of my absence, as well as providing me with some of the early interviews to help begin developing a rating scale. In particular it was Mary Dunn who several years ago encouraged me to think that this was a project I could usefully attempt.

In 1999 I was awarded a Travelling Fellowship by the Winston Churchill Memorial Trust that enabled me to travel to Israel and meet many people who were actively using the 6PSM in different settings. This was not only an invaluable information-gathering opportunity, but also the first time I had been awarded substantial funds in a competitive application. Without this I would never have had the confidence to apply for the NHS Fellowship, so I am deeply grateful for the opportunity the Churchill Trust provided.

Tracey Dent-Brown has been an invaluable sounding-board throughout the four years of the research, in particular at the stage of inter-rater reliability testing. In the final stages she and Dr Lewis Pike have been very helpful in providing feedback on the thesis chapters.

Dr Eric Gardner of the Clinical Psychology Department, University of Hull, gave some very valuable feedback and advice on the first draft of this thesis and suggested some helpful alterations and corrections.

Finally, as always, I want to thank my family Tracey, Rosie and Ella who have provided constant, unflagging love, warmth, support and cups of tea over the last four years. Now you can have the study back!

3.0 CONTENTS

1.0 ABSTRACT.....	2
2.0 ACKNOWLEDGEMENTS.....	3
3.0 CONTENTS.....	6
4.1 INTRODUCTION	16
4.2 LITERATURE REVIEW	18
4.2.1 THE PROJECTIVE HYPOTHESIS.....	18
4.2.1.1 <i>Criticisms of the projective approach</i>	21
4.2.2 THE SIX-PART STORY METHOD	24
4.2.2.1 <i>Description of the 6PSM</i>	24
4.2.2.2 <i>Existing literature on the 6PSM</i>	25
4.2.2.3 <i>Origins and development of the 6PSM: Morphological and semiological studies</i>	27
4.2.2.4 <i>Alida Gersie and the development of Story Evocation Techniques</i>	30
4.2.3 PERSONALITY DISORDER	33
4.2.3.1 <i>The development of the concept of personality disorder</i>	33
4.2.3.2 <i>Conceptions of personality and personality disorder</i>	37
4.2.3.3 <i>Categorical versus dimensional systems of personality disorder classification</i>	40
4.2.3.4 <i>Summary of position on personality disorder</i>	41
4.2.4 THE 6PSM IN PERSONALITY DISORDER	42
4.2.5 CONCEPTS OF RELIABILITY AND VALIDITY	46
4.2.5.1 <i>What would be the features of a reliable 6PSM?</i>	47
4.2.5.2 <i>What would be the features of a valid 6PSM?</i>	48

5.0 PLANNING OF STUDY, METHODS AND ANALYSIS.....	50
5.1 OUTLINE RESEARCH QUESTIONS	51
5.1.1 INTER-RATER RELIABILITY	51
5.1.2 TEST-RETEST RELIABILITY.....	52
5.1.3 RATER BIAS	54
5.1.4 CONCURRENT VALIDITY	54
5.1.5 PARTICIPANTS' EXPERIENCE OF THE 6PSM.....	55
5.2 METHODOLOGICAL AND EPISTEMOLOGICAL CONSIDERATIONS.....	56
5.2.1 PREFERRED METHODOLOGIES IN DRAMATHERAPY RESEARCH	56
5.2.2 QUALITATIVE OR QUANTITATIVE APPROACHES?	59
5.2.3 EPISTEMOLOGICAL POSITION	60
5.2.4 RATIONALE FOR SELECTION OF A MIXED METHODOLOGY.....	61
5.3 LESSONS FROM PUBLISHED STUDIES	64
5.3.1 QUANTITATIVE STUDIES.....	64
5.3.1.1 <i>Projective storymaking approaches to personality</i>	64
5.3.1.2 <i>Validity and reliability testing of existing storymaking assessments</i>	66
5.3.2 QUALITATIVE STUDIES.....	70
5.4 OUTLINE PLAN OF RESEARCH PROGRAMME	73
5.4.1 PLANNING OF PILOT STUDY.....	73
5.4.2 RECRUITMENT OF PARTICIPANTS	74
5.4.3 DATA TO BE COLLECTED.....	75
5.4.4 CONCURRENT MEASURES TO BE USED	75

5.4.4 CONCURRENT MEASURES TO BE USED	75
5.4.4.1 SCID-II	76
5.4.4.2 IIP-32	78
5.4.4.3 CORE-OM	80
5.4.5 SUMMARY PLAN FOR QUANTITATIVE ARM OF STUDY	81
5.4.6 QUALITATIVE ARM OF STUDY	82
6.0: PILOT STUDY	84
6.1 PLACE OF THE PILOT WITHIN THE STUDY AS A WHOLE.....	85
6.2 AIM OF PILOT STUDY	86
6.3 DEVELOPMENT AND METHOD OF PILOTING	86
6.3.1 SELECTION OF STORIES FOR PILOTING	86
6.3.2 PILOT RESPONSE FORM.....	89
6.3.2.1 <i>Quantitative Data</i>	89
6.3.2.2 <i>Categorical and Qualitative Data</i>	93
6.3.2.3 <i>Data on Raters</i>	96
6.3.2.4 <i>Recruitment of Pilot Raters</i>	96
6.4 RESULTS OF PILOT STUDY.....	97
6.4.1 PROFESSION, EXPERIENCE AND CONFIDENCE.....	97
6.4.2 PRELIMINARY ANALYSIS OF NUMERICAL DATA	98
6.4.3 PRELIMINARY ANALYSIS OF CATEGORICAL DATA.....	103
6.4.4 ANALYSIS OF RELATIONSHIPS FROM GRIDS	104
7.0: INTRODUCTION TO THE MAIN STUDY	106
7.1 OVERALL STUDY DESIGN.....	107

7.2 ESTABLISHING A PROTOCOL FOR ADMINISTERING THE	108
6-PART STORY METHOD	108
7.3 RECRUITMENT OF PARTICIPANTS	109
7.4 ETHICAL APPROVAL.....	110
7.5 TRAINING OF CLINICIANS	110
7.6 SELECTION OF PATIENTS	111
7.7 DATA GATHERING	115
7.7.1 DATA FROM FIRST STORY SESSION.....	115
7.7.2 DATA FROM INTERVIEW WITH RESEARCHER.....	115
7.7.3 DATA FROM SECOND STORY SESSION.....	117
7.8 RECRUITMENT AND TRAINING OF ADDITIONAL RATERS	117
8.0 PRIMARY DATA: COLLECTION AND PROCESSING.....	119
8.1 INITIAL PROCESSING OF DATA	120
8.1.1 PICTURES	120
8.1.2 TAPES.....	120
8.1.3 SCID-II INTERVIEW	123
8.1.4 CORE AND IIP-32 QUESTIONNAIRES	123
8.1.5 OTHER DATA	124
8.2 CHARACTERISTICS OF THE SAMPLE.....	124
9.0 SECONDARY DATA:	131

REFINEMENT AND COLLECTION	131
9.1 OBJECTIVES AND DESIGN	132
9.2 METHOD	133
9.2.1 ITEM SUITABILITY AND INTER-RATER RELIABILITY.....	137
9.2.2 FIRST ASSESSMENT OF ITEM INTER-RATER RELIABILITY	138
10.0: RATER ANALYSIS OF STORIES: RESULTS	142
10.1 STRUCTURE OF CHAPTER.....	143
10.2 INTER-RATER RELIABILITY OF INDIVIDUAL ITEMS.....	143
10.3 CONSTRUCTION OF SCALES	146
10.3.1 SCALE ANALYSIS OF REMAINING ITEMS	146
10.3.2 RELIABILITY OF SCALES	148
10.4 STORIES FROM PATIENTS WITH OTHER PD DIAGNOSES	154
10.4.1 SCHIZOID PERSONALITY DISORDER	155
10.4.2 AVOIDANT PERSONALITY DISORDER	156
10.5 CORE CONFLICTUAL RELATIONS THEME (CCRT) RATINGS.....	157
10.6 RATER INFERENCE OF GLOBAL MENTAL HEALTH	160
10.7 RATER INFERENCE OF AUTHOR GENDER.....	163
10.8 ASSOCIATION OF RATER MEASURES WITH	164
OTHER FEATURES OF STORY DATA	164
10.9 THE EFFECT OF RATER CHARACTERISTICS ON STORY RATINGS	165

10.9.1 ASSESSING BIAS IN RATINGS OF STORIES	165
10.9.2 INFLUENCE OF RATER EXPERIENCE	167
11.0 RATER ANALYSIS OF STORIES: DISCUSSION	169
11.2 SCALE ASSEMBLY FROM STATEMENTS	171
11.3 PSYCHOMETRIC PROPERTIES OF NEGATIVITY SCALE	176
11.4 INFLUENCE OF AUTHOR GENDER	177
11.5 CHARACTERISTICS OF STORIES FROM PATIENTS WITH	178
SCHIZOID PERSONALITY DISORDER.....	178
11.6 CHARACTERISTICS OF STORIES FROM PATIENTS WITH	179
AVOIDANT PERSONALITY DISORDER.....	179
11.7 CORE CONFLICTUAL RELATIONS THEME (CCRT) RATINGS.....	180
11.8 RATER INFERENCE OF GLOBAL MENTAL HEALTH	180
11.9 COMPARISON OF MORE EXPERIENCED AND.....	182
LESS EXPERIENCED RATERS.....	182
12.0 QUANTITATIVE TEXTUAL ANALYSIS	184
12.1 INTRODUCTION	185
12.2 INVESTIGATIONS OF TRANSCRIPT LENGTH.....	185
12.2.1 STABILITY	185
12.2.2 VARIATIONS IN STORY LENGTH	186

12.3 COMPUTER BASED TEXT ANALYSIS.....	193
12.3.1 GENERAL INQUIRER.....	193
12.3.1.1 <i>Constructing a measure of borderline personality disturbance from GI data</i>	194
12.3.1.2 <i>Comparison of GI-B score with concurrent measures</i>	200
12.3.2 LINGUISTIC INQUIRY AND WORD COUNT (LIWC).....	203
12.3.2.1 <i>Correlation with concurrent measures</i>	206
12.4 THE REGRESSIVE IMAGERY DICTIONARY	207
12.5 CONSTRUCTION OF NEW CATEGORIES	211
12.6 SUMMARY OF RESULTS OF COMPUTER BASED TEXTUAL ANALYSIS	215
12.7 QUANTITATIVE TEXT ANALYSIS: OTHER FEATURES	216
13.0 QUANTITATIVE TEXTUAL ANALYSIS:.....	220
DISCUSSION.....	220
13.1 DISCUSSION OF STORY LENGTH.....	221
13.2 COMPUTERISED TEXT ANALYSIS.....	223
13.3 CUSTOM-BUILT DICTIONARY FOR BASICPh	230
13.4 OTHER FEATURES OF QUANTITATIVE TEXT ANALYSIS.....	231
14.0 QUALITATIVE ANALYSIS: METHOD AND RESULTS	235
14.1 FOREWORD TO THIS CHAPTER	236

14.2 AIMS OF THE QUALITATIVE ANALYSIS	237
14.3 METHOD	237
14.3.1 DATA COLLECTION AND PREPARATION.....	237
14.3.2 DATA ANALYSIS	239
14.3.3 CODING.....	241
14.3.3.1 <i>Identification of concepts</i>	241
14.3.3.2 <i>Amalgamation of concepts into categories</i>	243
14.3.3.3 <i>Finding properties and dimensions of categories</i>	244
14.3.3.4 <i>Axial coding</i>	244
14.4 RESULTS	246
14.4.1 IDENTIFICATION OF CATEGORIES.....	246
14.4.1.1 <i>First category: Parallels between story and own situation</i>	246
14.4.1.2 <i>Second category: Progressive development of story</i>	249
14.4.1.3 <i>Third category: Response to process</i>	250
14.4.1.4 <i>Fourth category: Comparisons between first and second stories</i>	253
14.4.1.5 <i>Fifth category: Aesthetic distancing</i>	254
14.4.1.6 <i>Sixth category: Material evoked</i>	256
14.4.2 AXIAL CODING	258
14.4.3 IDENTIFYING THE CENTRAL CATEGORY.....	260
15.0 QUALITATIVE ANALYSIS: DISCUSSION.....	266
15.1 ACCEPTABILITY OF THE 6-PART STORY METHOD.....	267
15.2 A THEORY FOR THE OPERATION OF THE 6PSM	271
15.2.1 GROUNDED THEORY AND THE CENTRAL CATEGORY	273

15.2.2 EXISTING MODELS FOR UNDERSTANDING THE ACTION OF THE 6PSM.....	275
15.2.3 CONSTRUCTING AN UNDERSTANDING - A COGNITIVE MODEL	278
16.0 FINAL DISCUSSION	286
16.1 INTRODUCTION AND REVIEW OF RESULTS	287
16.1.1 CAN A PRACTICAL RATING TECHNIQUE FOR SIX-PART STORIES BE DEvised?.....	288
16.1.2 CAN TWO RATERS INDEPENDENTLY RATE STORIES RELIABLY?	289
16.1.3 ARE RATINGS OF DIFFERENT STORIES FROM THE SAME PARTICIPANTS STABLE OVER TIME?	291
16.1.4 DO THE RATINGS GIVEN BY RATERS TO OTHERS' STORIES BEAR ANY RELATION TO THE PERSONAL CHARACTERISTICS OF THE RATER THEMSELVES?	293
16.1.5 ARE THE RESULTS FROM THE 6PSM VALIDATED BY ANY OF THE CONCURRENT DATA GATHERED ABOUT PARTICIPANTS?	294
16.1.6 IS THE 6PSM ACCEPTABLE TO PATIENTS AS A METHOD FOR GENERAL CLINICAL USE?.....	296
16.1.7 DOES THE 6PSM PRODUCE MATERIAL THAT HAS FACE VALIDITY FOR THE AUTHORS THEMSELVES?	297
16.1.8 DO THE PARTICIPANTS' ACCOUNTS PERMIT THE DEVELOPMENT OF A THEORY ABOUT HOW THE 6PSM WORKS?.....	297
16.2 RELATION TO PREVIOUSLY PUBLISHED RESULTS.....	298

16.3 LIMITATIONS OF PRESENT STUDY.....	300
16.4 IMPLICATIONS FOR CLINICAL PRACTICE	304
16.4.1 TRAINING IMPLICATIONS	304
16.4.2.1 <i>What do the idiographic data reveal about the borderline situation?</i>	307
16.5 IMPLICATIONS FOR FURTHER RESEARCH	308
16.5.1 CONFIRMATION OF VALIDITY	308
16.5.2 EXTENSION OF NORMATIVE DATA.....	308
16.5.3 ASSESSMENT OF RATER BIAS.....	309
16.5.4 ANALYSIS BY AUTHORS.....	310
16.5.5 EFFECTS ON ALLIANCE	310
16.5.6 EFFECTS ON THERAPEUTIC OUTCOME	311
17.0: APPENDICES	313
APPENDIX 1: PILOT RESPONSE FORM	314
APPENDIX 2: PATIENT INFORMATION LETTER AND CONSENT FORM...325	
APPENDIX 3: PROTOCOL FOR ELICITING 6-PART STORY MATERIAL328	
APPENDIX 4: INSTRUCTIONS TO FINAL RATERS AND FINAL RATING FORM	335
APPENDIX 5: RESULTS OF MINOR ANALYSES (SEE SECTION 10.8)	339
18.0 BIBLIOGRAPHY.....	341

4.0 INTRODUCTION AND LITERATURE REVIEW

4.1 INTRODUCTION

Projective techniques are arguably central to the professions of dramatherapy and the other arts therapies. All these disciplines rely on the production and exploration of *projected pieces of drama, art, music or movement*. And yet there has been little or no empirical research within the arts therapies to establish the evidence base for these techniques. A recent review of assessment and outcome measures used in the arts therapies (Kleindienst & Frude, 1999) reported finding no empirically validated tools, projective or otherwise, in the arts therapies research literature.

Conversely, projective techniques remain a small but vigorously discussed part of the practice of psychology and psychotherapy. Regardless of criticisms (Lilienfeld, Wood, & Garb, 2000), the literature describing such techniques continues to grow, as do the numbers of new instruments (Costantino & Malgady, 1999; Coulacoglou & Kine, 1995; Edwards, 1996), alongside better known techniques such as the revision of the Rorschach system (Exner, 1995).

This study is an investigation of one projective assessment tool well known to dramatherapists and others. The 6-Part Story Method (6PSM) of assessment was chosen because it is widely taught on UK pre-registration courses for dramatherapists and is frequently referred to in literature on assessment in

dramatherapy (Forrester, 2000; Pendzik, 2003). It is also use in clinical practice by the NHS psychotherapy service in which the author works (Dent-Brown, 1999a; Dunn & Parry, 1997). This service works principally with patients who have a diagnosis of personality disorder.

The 6PSM involves the clinician giving a structured set of instructions to the patient, following which the patient creates, recounts and discusses a new, fictional story. The assumption is that the relationships, world-view, problems and coping strategies described in the story bear some relation to those elements in the life of the patient who has devised it. In the clinical setting the 6PSM very often appears to bring up useful and interesting material, but despite the popularity and widespread use of the 6PSM no systematic investigation of the 6PSM appears to have been carried out to date. The question of the validity and reliability of the clinical material results has never been addressed, nor has any formal investigation of the mode of action of the 6PSM been undertaken. This study is an attempt to answer some of those questions.

4.2 LITERATURE REVIEW

4.2.1 THE PROJECTIVE HYPOTHESIS

It is not a recent discovery that stimuli of all sorts can have other images projected onto them:

HAMLET: Do you see yonder cloud that's almost in shape of a camel?

POLONIUS: By th' mass and 'tis - like a camel indeed.

HAMLET: Me thinks it is like a weasel.

POLONIUS: It is backed like a weasel.

(Hamlet, Act III, Sc 2)

In a similar vein, but with a different stimulus, there was a popular parlour game in late 19th century Germany, comprising of a series of ambiguous inkblots. The game was known to one Hermann Rorschach, a doctoral student of Eugen Bleuler. Rorschach noticed that schizophrenic patients seemed to give very different responses to the inkblot stimulus. He published his findings (Rorschach, 1921) and so the Rorschach system (Exner, 1995) was born. Although not a storymaking tool, the Rorschach system gave birth to other approaches that produced more recognisable story material such as the Thematic Apperception Test (TAT) (Murray, 1943; Stein, 1955).

Rorschach did not advance a theoretical hypothesis for his findings, but this was provided in the following decade by Frank (1939). It was Frank who first suggested that this tendency of the human brain to perceive meaningful patterns in relatively meaningless stimuli might be useful in the assessment of

personality. He called this the projective hypothesis for the study of personality. He was concerned that in conventional, self-report methods of assessment, subjects may not have the insight or psychological-mindedness to be able to describe themselves accurately; or if they were able they might be unwilling to do so because of a need to give socially acceptable answers:

“...many procedures for study of personality rely upon the subject's self-diagnosis and revelation of his *private world* of personal meanings and feelings which the social situation compels the individual to conceal, even if, as is unusual, he had any clear understanding of himself.”

(Frank, 1939: p395, italics in original)

Frank used the physical sciences to provide a metaphor for phenomena and structures which cannot be directly observed. The configuration of atoms or the structure of crystal lattices must be inferred from indirect data such as the diffraction of light or the behaviour of other substances. An electron can never be seen directly, but its path can be traced in a cloud chamber by observing its interaction with other particles. Thus the idea of the more indirect projective technique:

“In similar fashion we may approach the personality and induce the individual to reveal his way of organising experience by giving him a field (objects, materials, experiences) with relatively little structure and cultural patterning..... thus we elicit a projection of the individual personality's *private world* because he has to organise the field, interpret the material and react affectively towards it.” (Frank, 1939 pp.402-3, italics in original.)

Of course Frank was not the originator of the idea of projection. The concept was enunciated 28 years earlier by Freud (1911) for whom it was a defence mechanism whereby individuals unconsciously identified characteristics and behaviour in others that were actually their own, unacceptable traits. This concept of "classical projection" remains current, although it has not apparently been susceptible to experimental confirmation (Holmes, 1974). But as Lilienfeld, Wood & Garb (2000) point out, projective assessment techniques draw on a rather different form of projection. What they characterise as generalised or assimilative projection is

"... the relatively uncontroversial tendency for individuals' personality characteristics, needs and life experiences to influence their interpretation ('apperception') of ambiguous stimuli."

(Lilienfeld *et al.*, 2000: p29)

The kinds of projective approaches referred to by Frank include the Rorschach test as well as puppetry, the making of music and art, block play, story completion and dramatic role play. He seemed to suggest that these approaches could bring out the individual uniqueness of the subject, in contrast to the way conventional psychometric tests revealed their similarity to others. In talking this way Frank set projective and psychometric approaches as polar opposites (with the former being implicitly privileged, in his view.) It appeared to follow that the statistical tools for ensuring reliability and validity of psychometric tests were neither helpful nor even necessary for investigating projective techniques. Of projective methods of measurement he says "... these indirect methods rely upon experimental and genetic procedures to establish validity and reliability, not statistical procedures." (Frank, 1939: p399)

4.2.1.1 Criticisms of the projective approach

This perceived antagonism between projective and psychometric tests may have led to the situation described by Masling:

“The early years of using this test were marked by the naïve belief that the Rorschach method properly applied could help reveal the mysteries of the human psyche. Such considerations as reliability, validity, and the influence of situation and examiner were scarcely raised, let alone answered.” (Masling, 1997: p258)

This dearth of data, and the perceived antipathy of proponents of projective techniques to rigorous statistical analysis has led to a fall from grace for projective techniques. From being pre-eminent in the 1940s and 1950s, the Rorschach became an almost shameful anachronism. Jensen’s famous quote is only the most trenchant of many:

“The rate of scientific progress in clinical psychology might well be measured by the speed and thoroughness with which it gets over the Rorschach.” (Jensen, 1965: p238)

Recently a more balanced view has been evident, with proponents of projective methods being more modest in their claims (and more methodical in their justification of them). Meanwhile critics have become less strident, criticising excesses where practice is not based on evidence, while accepting that there is a sound theoretical base for projective techniques. For example, while being

highly critical of the way projective techniques have been used in clinical practice, Lilienfeld *et al.* acknowledge that:

“...it is evident that certain projective instruments, as well as scores derived from these measures, can indeed achieve acceptable levels of reliability and validity. Consequently, dismissing in broad brush all projective techniques as unreliable and invalid is unwarranted.” (Lilienfeld *et al.*, 2000: p53)

At the same time, a defender of the Rorschach is happy to observe that:

“...the hope that projective tests would function like psychic X-rays, revealing the personality structure of the patient in an antiseptic environment free of influence from any source other than the patient’s psyche proved to be both naïve and illusory.” (Masling, 1997: p259)

One particular problem that Masling identifies is that in interpreting a projective session the projections identified are just as likely to be those of the clinician as those of the patient. A further complicating problem is that of the Barnum effect (Dickson & Kelly, 1985). This refers to the phenomenon where, following some form of personality assessment, diffuse, vague interpretive statements are made about the subject which could in fact apply to almost anyone. An example would be “While you have some personality weaknesses you are generally able to compensate for them.” The phenomenon is named after the American showman P.T. Barnum who attributed his success to having a little something for everybody in his show. As Snyder and Shenkel (1975) point out he also said “There’s a sucker born every minute.”

Further investigations of the Barnum effect have demonstrated that the results of projective techniques are particularly likely to be accepted as true by subjects, more so than the results of psychometric tests (Hanson, Claiborn, & Kerr, 1997). Furthermore, the more complex the assessment procedure and the more information is asked for, the more the result is rated by the subject as being accurate. Projective techniques, with their high levels of interaction and the possibility of magical thinking about what one is revealing may be particularly susceptible to this problem (Snyder & Shenkel, 1976).

In summary, projective tests have been a staple of psychological testing in the past, particularly in the USA. Excessive, unjustified claims about them led to a fall from grace and the rise of personality assessments based on actuarial methods with demonstrable statistical reliability and validity such as the MMPI. As a result, there is a very full psychological literature about projective tests, but in practice very few psychologists (in the UK at least) use them routinely. By contrast, dramatherapists place projective techniques as central to their work, and yet there has been no formal research and little published description of these techniques. This is not to be wondered at in a profession which is still very young and small in numbers, but remains a gap to be filled.

4.2.2 THE SIX-PART STORY METHOD

4.2.2.1 *Description of the 6PSM*

The application of the 6PSM has been described in full by its originator (Lahad, 1992). The instructions for administering the 6PSM as used in this study are at Appendix 3. To summarise the procedure, the patient or participant is given a blank sheet of paper and a pen or pencil. They are asked to divide the paper into six spaces to work in. The patient then follows instructions to fill in the spaces with images to represent the following (in Lahad's original format):

1. A main character in the setting where the story begins
2. A problem facing the main character in the story
3. Things that help the main character to cope with the problem
4. Things that get in the way of the main character
5. The main action, how the main character copes with the problem
6. What happens after the main action

Once the six pictures are drawn the patient is asked to tell the story, without interruption or questions. They are to tell it in as full and detailed a way as possible, adding detail and inventing new descriptions as they go. Finally the clinician or researcher asks questions about each picture, and the story in general, to elaborate the story and check any points that are not clear.

4.2.2.2 Existing literature on the 6PSM

The first mention of the 6PSM in print was by the originator Mooli Lahad (1992).

This book chapter describes the process of eliciting the story outlined above, then goes on to discuss a method of analysing the story known as BASIC Ph.

This acronym stands for the six methods of coping which Lahad felt could be identified in 6-part stories; methods based on Belief, Affect, Social, Imaginative, Cognitive or Physical coping. A neat distinction between the 6PSM and BASIC Ph has been made by Pendzik (2003):

"We may call *technique* or *tools* the means or instruments utilized in order to generate data and the *method* or *approach* the criteria by which the data is [*sic*] analysed....Thus for instance, a story created through 6PSM may also be assessed through Landy's role method; by the same token, Johnson's Role-Playing Test can be read in terms of the BASIC Ph theory." (Pendzik, 2003: p93, italics in original)

Lahad's second publication (Lahad & Ayalon, 1993), also a book chapter, was similar in content to the first, adding a number of examples of stories elicited using the 6PSM and using BASIC Ph to try to understand them. Both these two chapters concentrate on using the 6PSM to identify preferred methods of coping.

The two publications by Lahad make implicit claims about the validity of the method in several places:



“My assumption is that by telling a projected story based on the elements of fairytale or myth, I may be able to see the way the self projects itself in organised reality in order to meet the world.” (Lahad, 1992: p157)

“So, it seems that with the aid of the structured story, a person’s coping resources and conflict areas can be located relatively quickly.” (Lahad & Ayalon, 1993: p18)

Lahad does however note the many criticisms levelled against projective techniques, listing seven areas of concern including the lack of standardised administration instructions and the concerns about low validity and reliability.

However he says:

“Most of the above [concerns] are less evident in the 6PSM because of its nature and the way it is administered. Reliability is problematic, whether inter-measurement (i.e. between projective techniques of other kinds) or with different judges.” (Lahad & Ayalon, 1993: p24)

Since the publication of these descriptions of the 6PSM, the method has become one of the standard methods of assessment in dramatherapy, often referred to in the professional literature. For example Landy *et al.* (2003) make reference to both 6PSM and BASIC Ph in their literature review of dramatherapy assessment instruments, as does Pendzik (2003). However neither author, nor Lahad himself, makes any reference to studies of the reliability and validity of the method. Forrester (2000) does refer to dramatherapy assessments that have undergone validity testing, such as the

Diagnostic Role-Playing Test (Johnson & Quinlan, 1993), but the 6PSM is not among those assessments discussed.

Subsequent published accounts of the use of the 6PSM (Dent-Brown, 1999a, 1999b, 2001b) have described its use in an NHS personality disorder service.

These articles take for granted that the data produced by the 6PSM can be relied upon as a replicable and valid indicator of the story-teller's personality.

This assumption may be necessary in the building of a technique, which must be developed and found to be clinically feasible and useful in the first place. But although necessary, this assumption alone cannot be sufficient if a technique is to be regularly used for clinical decision making. It was for this reason that this reliability and validity study of the 6PSM was planned.

4.2.2.3 Origins and development of the 6PSM: Morphological and semiological studies

The 6PSM has its roots in the early 20th century morphological study of fairy tales and the later semiological studies that followed. In both disciplines the search was for general, universal factors that were common to particular, individual stories.

Early in the 20th century the greatest contribution came from Vladimir Propp (1968) whose study The Morphology of the Folktale was originally published in Russian in 1928. The title is slightly misleading, as all the 270 tales described are in fact fairy tales, not folktales. Nevertheless, this work is the foundation

stone for modern structural story analysis, both in the pure field of semiotics and in the applied fields of the psychotherapies.

Propp was interested in common themes running through the extensive canon of Russian fairy tales, and he produced a list of *dramatis personae* and elements that he felt were exhaustive. Although neither every actor nor every element appeared in every story, he believed that he had identified a sequence of events and characters that always appeared in a certain order. He made four observations that summarise his work (Propp, 1968: p23):

- “1. Functions of characters serve as stable, constant elements in a tale, independent of how and by whom they are fulfilled. They constitute the fundamental elements of a tale.
2. The number of functions known to (*sic*) the fairy tale is limited.
3. The sequence of functions is always identical.
4. All fairy tales are of one type in regard to their structure.”

The *dramatis personae* he identified were these:

- The hero
- The dispatcher (who gives the hero the task)
- The princess sought by the hero
- The princess' father
- The villain (who opposes the hero)
- The provider (who gives things that help the hero)
- The helper (who actively aids the hero)
- The anti-hero (who impersonates the hero or tries to steal his prize)

Propp was not saying that every story had exactly these eight characters; in a particular story the villain and the anti-hero might be the same person, and the dispatcher may also be the provider for example. But he maintained that all these functions were played out in every Russian fairy tale he analysed.

So far this analysis was restricted to the very circumscribed genre of Russian fairy tales. In the 1950s French structuralists and semioticians took great interest in Propp's work, starting with Lucien Tesnière (1959) who looked at the *dramatis personae* and came up with the concept of the actant. He defined actants as:

"...beings or things that participate in the process (of the story) in any way whatsoever, even as mere walk-on parts or in the most passive way.¹"

This helpful definition moves the focus wider than just people. Tesnière makes it clear that animals and even inanimate objects can be actants; for example a story about a prisoner in a cell seems only to have one actor, the prisoner struggling for freedom. But there are two actants; the cell that confines the prisoner is just as much a part of the story as the prisoner him or herself.

¹ "...les êtres ou les choses qui, à un titre quelconque et de quelque façon que ce soit, même au titre de simples figurants et de la façon la plus passive, participent au procès."

Subsequently Algirdas Greimas (1966) used Tesnière's concept of actants to codify Propp's *dramatis personae*, simplifying them into a system of six actants set out in Figure 4.1 below:

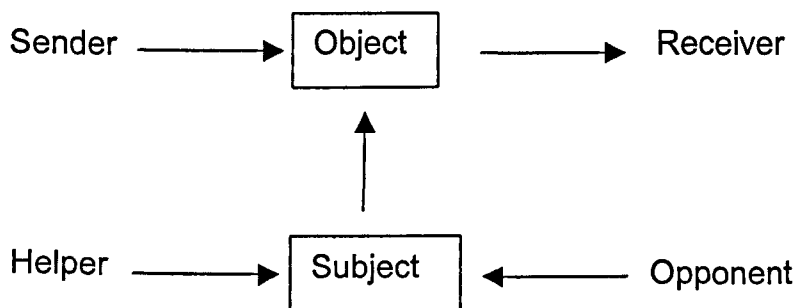


Figure 4.1: Functional organisation of Greimas's six actants

Greimas considered that this structure might describe all stories, not just Propp's large but specialised body of fairy stories. The core of the story is the subject-object pair, or what might be seen as the hero and their task.

4.2.2.4 *Alida Gersie and the development of Story Evocation Techniques*

In his introduction to the 2nd edition of Propp's *Morphology*, Alan Dundes makes some suggestions (p.xv) about the implications of Propp's (and by extension Greimas's) findings. He suggests the construction of story stems as prompts to see how children respond, saying that "such a test might also be of value in studies of child psychology". He also suggested that "Propp's scheme could also be used to generate new tales"; a suggestion noted and taken up subsequently by Alida Gersie.

Gersie is an Anglo-Dutch dramatherapist who has published extensively on the therapeutic use of stories (Gersie, 1991, 1992, 1997; Gersie & King, 1990) and who has been teaching and supervising dramatherapists and others in the use of story since the late 1970s. She developed methods that helped a client create a new story which she called Story Evocation Techniques (SETs). These were question-based techniques where the participant would be asked open questions to establish the framework of the story on which they would then elaborate. These SETs were taught to other dramatherapists but had not been written about in detail until more recently (Gersie, 2002, 2003a, 2003b).

Two of the therapists who were taught by Gersie in the early 1980s were Mooli Lahad and Ofra Ayalon. Lahad was an educational psychologist and dramatherapist working in the educational system in northern Israel, while Ayalon was a child and adolescent psychotherapist working nearby. Lahad worked in the northern Galilee panhandle, a narrow strip of Israeli territory that was at that time subject to frequent mortar, artillery and rocket attacks as well as cross-border incursions from neighbouring Lebanon and Syria. He was interested in working with schools to develop pupils' resilience to trauma, and developed the SET he had been taught into the 6-part story method (Lahad, personal communication, 15/11/99). The point of the 6PSM was not that the storymaking should in itself be therapeutic, but that the coping elements in the story might give insight into the preferred coping strategies of their authors (Lahad, 1992; Lahad & Ayalon, 1993). The 6PSM has since become widely used in Israel, in settings and for purposes as diverse as personnel selection in education and to monitor the emotional wellbeing of children in a paediatric oncology ward (Dent-Brown, 2001a).

In the 1990s Lahad travelled frequently to the UK to deliver training, and the 6PSM was taught by him to a new generation of dramatherapists and others. Subsequently it has become a standard part of the syllabus of pre-registration dramatherapy courses in the UK. The 6PSM was adopted as part of a wider patient assessment package in an NHS personality disorder service (Dunn & Parry, 1997) with a view to the 6PSM triangulating with self-report and clinical interview to maximise the information gained from patients (Dent-Brown, 1999a). This use of the 6PSM has generated much interest (Dent-Brown, 1999b) but a lengthy search found no publications reporting the results of empirical research into the method.

4.2.3 PERSONALITY DISORDER

The 6PSM could not be studied in a vacuum, the stories must come from some participant group in a recognisable context. The author's clinical team was a tertiary psychotherapy service supporting secondary community mental health teams in their work with patients with a personality disorder. This was the context and the participant group chosen; partly for convenience, partly because there was already some clinical experience in using the 6PSM here, and partly because there would be some immediate results of clinical use if the study was completed.

4.2.3.1 *The development of the concept of personality disorder*

Individuals who retained their intellectual capacity but harboured strange and unrealistic ideas had puzzled European physicians of the 18th century. Philippe Pinel (1745-1826) termed the condition *manie sans délire*. The following century James Prichard (1835), a member of the Parliamentary Lunacy Commission coined the term 'moral insanity', by which he meant:

"a morbid perversion of the feelings, affections and active powers, without any illusion or erroneous conviction impressed on the understanding: it sometimes coexists with an apparently unimpaired state of the intellectual faculties." (Prichard, 1835: p 14)

He went on to paint a picture which is still recognisable today of people who

"...think and act under the influence of strongly-excited feelings, and persons accounted sane are, under such circumstances, proverbially liable to error both in judgement and conduct) (Prichard, 1835: p15)

This is arguably the beginning of the identification of disorders of personality or character as somehow separate from other classes of mental health problems. The term 'moral insanity' continued to be used throughout the 19th century – the assassin of US President Garfield in 1888 used it as a defence (Ozarin, 2001) – but towards the end of the century it began to be more diversely defined. One strand saw the more antisocial features reclassified as 'psychopathic inferiority' by Koch in Germany who described patients "who conform to a certain intellectual standard but who throughout their lives exhibit disorders of conduct of an antisocial or asocial nature" (Ozarin, 2001).

A second strand concentrated on those who were more of a risk to themselves than others; one author (Rosse, 1890) coined the term 'borderland insanity' for patients with 'neuropathic' tendencies (urges towards suicide or other self-damaging behaviour.) This may be the origin of the modern term borderline, although nobody was yet using the term personality disorder.

By the 1920s the development of a psychoanalytic language had allowed the elaboration of the concept of the borderline between psychosis and neurosis. Patients on this borderline showed some, but not all, of the features of both states (Stern, 1938) and while the clinical picture resembles very closely that of the modern diagnosis of BPD they were seen as suffering from an attenuated form of schizophrenia. Stern's description included ten features:

“(1) narcissism, (2) paralysis in the face of crises (“psychic bleeding”), (3) inordinate hypersensitivity, (4) bodily and physical “rigidity”, (5) negative therapeutic reaction, (6) constitutional feeling of inferiority, (7) masochism, (8) “organic insecurity”, (9) projective mechanisms and (10) difficulties in reality testing.” (Stone, 1986: p478)

Thirty years later (and after the discovery of the neuroleptics) there had been a further divergence and BPD was no longer seen in this way. Kernberg (1967) outlined his description of the Borderline Personality Organisation, a description which strongly influenced the description of Borderline Personality Disorder when it first appeared in the DSM-III (American Psychiatric Association, 1980: p23). There it has stayed ever since, regardless of the criticisms of its status, which will now be outlined.

Arntz (1999) summarises some of the objections to the concept of personality disorder in general and BPD in particular. First there are theoretical objections that spring from the psychoanalytic origins of the concept of personality disorder: theoreticians, clinicians and researchers from other orientations such as cognitive or behavioural schools have rejected the concept as relying too heavily on psychodynamic assumptions. This ignores the fact that behaviourally-oriented authorities such as Millon (1981) have had a major role in developing the DSM definitions of personality disorder. In addition, there are counter-objections from more psychodynamically-oriented clinicians who have criticised the DSM Axis II definitions for their emphasis on externally observable behaviours at the expense of subjective patient experience.

A second theoretical objection is the polythetic nature of DSM (and ICD) criteria; i.e. only a number of criteria for any disorder have to be met, leading to a bewildering number of variations of ways in which a diagnosis can be made. In what sense then are these diagnoses reporting on a single, central phenomenon? However this objection can be made of Axis I or physical diagnoses such as Rheumatoid Arthritis just as well, without leading to doubts about the fundamental existence of the disorders described. Moreover, some research has suggested that a core phenomenon is discernible among the criteria for BPD and that the diagnosis may be describing more than an imaginative construct (Burgmer, Jessen, & Freyberger, 2000; Sansone, Songer, & Gaither, 2001)

Three more practical objections have been raised to the DSM-defined personality disorders. The first relates to how far the disorders can be distinguished from one another, the criticism being that they overlap so far as to essentially indistinguishable. However Arntz (1999) reports on a confirmatory factor analysis undertaken on a sample of 404 patients which strongly supported the existing categorical structure of the DSM-III. Although small changes to some of the criteria were suggested (some of which have been implemented in DSM-IV) the existing set of PD diagnoses did in fact permit disorders and patients to be distinguished from one another.

The second practical objection relates to concurrent validity: do various measures of PD correlate or not, the criticism being that correlation is so poor that it is impossible to tell whether the fault lies with the measures or the fundamental definitions of the disorders. The same article by Arntz (1999)

mentioned above describes further data on diagnosis by SCID-II of a series of 317 patients. Correlation between the patient self-report questionnaire and the therapist interview were good, ranging from .57 to .99. However patients were then diagnosed by clinicians who were ignorant of the SCID-II diagnosis and agreement here was poor. Where Kappas could be calculated they were low (from -.01 to .33) with the exception of BPD (Kappa = .66). This is certainly an ongoing problem for the DSM, although BPD seems to be a positive exception; perhaps as one of the most common and problematic of the PD diagnoses it is more likely to be accurately diagnosed by real-world clinicians?

The third practical objection relates to the difficulty in distinguishing Axis I and Axis II disorders, the objection being that what are called personality disorders (e.g. Depressive Personality Disorder) are simply particular forms of an Axis I mental illness. For example, is BPD merely a chronic, severe form or end-state of PTSD? One study (Driessen *et al.*, 2002) suggested that while there are many overlaps in current presentation, aetiological factors for the two diagnoses differ considerably. Also, Arntz (1999) found that while BPD and PTSD are correlated, one is not a sub-set of the other. Although diagnoses overlap, there were sufficient patients in their sample with one diagnosis but not the other to suggest that they are distinct (if linked) entities.

4.2.3.2 Conceptions of personality and personality disorder

The assumptions made about personality disorder for the present study must be made explicit here. When the study was begun, the author's team was drawn from psychotherapists with different backgrounds but mostly with a humanistic

orientation (rather than, for example, psychoanalytic, cognitive or behavioural.) The prevailing model in the team was that what are called personality disorders are in fact creative adaptations to extraordinary or extreme early life experiences that have become habitual or embedded. By adulthood these adaptations may have served their purpose and become problems in their turn, but no alternative way of being is available to the patient.

This view was informally articulated and was derived from clinical practice and from a generally shared humanistic orientation, rather than from an evidence or theoretical base in any particular profession or therapeutic school. The team frequently made use of the work of Manfield (1992) in formulating patients' difficulties. Manfield discusses borderline, narcissistic and schizoid disorders from a psychodynamic perspective, having been himself influenced by the psychoanalytic work of Kernberg (1975; 1984).

The difficulty this left was that none of these approaches to personality disorders were easy to use in a research context. The individual formulations produced by the author's team tended not to dwell on diagnostics, but on the unique and personal features of the patient at hand. Manfield does describe borderline, narcissistic and schizoid disturbances, but in terms useful to clinicians rather than a researcher and Kernberg's work too was difficult to transpose into a UK community mental health setting; it necessitates a close and long familiarity with psychoanalytic thought and practice that would be unavailable in an NHS secondary care setting. None of these approaches obviously lent themselves to a quantitative study correlating 6-part stories with the presence or absence of a particular problem, with the logical necessity of

dichotomous, either-or classifications (or at least some kind of quantifiable continuum.)

The obvious alternative was to use the categorical classifications of the DSM-IV (American Psychiatric Association, 1994) or ICD-10 (World Health Organisation, 1995). One great danger of such a system is that it may appear to be describing an external, objective phenomenon that can positively be determined as being present or absent in any given case. In addition there is the problematic inclusion of personality disorder alongside other disorders with a much clearer disease model of aetiology and treatment. Speaking of mental health problems generally, Szasz (1991) says: "If mental illnesses are diseases, they are diseases of the brain, not the mind. If mental illnesses are names of behaviour, they are forms of behaviour, not diseases." (Szasz, 1991: p1574). This problem is even more marked with personality disorder; should it be seen as a diagnosis at all, occupying a category also populated by asthma and athlete's foot?

In this study it was decided to accept the construct of the DSM-IV diagnosis because it was a useful heuristic device, not because it was necessarily felt to describe some unique, externally existing entity. In fact the position was taken that a diagnosis of, say, Borderline Personality Disorder is an arbitrary construct of the researcher, rather than being an inherent quality of the participant. But the point is that it is an arbitrary construct that is well known to other clinicians and researchers, and which can be recognised with some degree of agreement by different observers of the same individual, whatever its ontological status.

It was considered whether non-diagnostic measures of personality, such as the MMPI or 16PF should be used as the concurrent measure. It was decided not to use these for two reasons. First, these are measures of personality, not personality disorder. Second, these measures are not widely known in NHS community mental health practice, however well known they are to psychology researchers. A DSM-IV or ICD-10 diagnosis would not need as much explanation to a front-line clinician as an MMPI result.

4.2.3.3 Categorical versus dimensional systems of personality disorder classification

Debate is ongoing about the two approaches. In clinical practice dimensional systems have been described as being more useful, as well as describing more variance in behaviour, disability and outcome (Rosenman, Korten, Medway, & Evans, 2003). Unfortunately, there are currently no obvious candidates for a dimensional system of personality disorder diagnosis, although these are being discussed (Westen, Heim, Morrison, Patterson, & Campbell, in press).

However there are some possibilities for flexibility within the DSM-IV and ICD-10. Both rely on enumerating a list of criteria, the presence of a certain number of which will determine a positive diagnosis. For example, the DSM-IV description of Borderline Personality Disorder is:

1. "A pattern of instability of interpersonal relationships, self-image, and affects, and marked impulsivity beginning by early adulthood and present in a variety of contexts, as indicated by five (or more) of the following:
2. Frantic attempts to avoid real or imagined abandonment. *Note:* Do not include suicidal or self-mutilating behaviour covered in criterion 5.

3. A pattern of unstable and intense interpersonal relationships characterised by alternating between extremes of idealisation and devaluation.
4. Identity disturbance: markedly and persistently unstable self-image or sense of self.
5. Impulsivity in at least two areas that are potentially self-damaging (e.g. spending, sex, substance abuse, reckless driving, binge eating.)
Note: Do not include suicidal or self-mutilating behaviour covered in criterion 5.
6. Recurrent suicidal behaviour, gestures or threats, or self-mutilating behaviour.
7. Affective instability due to a marked reactivity of mood (e.g. intense episodic dysphoria, irritability or anxiety usually lasting a few hours and only rarely more than a few days.)
8. Chronic feelings of emptiness.
9. Inappropriate, intense anger or difficulty controlling anger (e.g. frequent displays of temper, constant anger, recurrent physical fights).
10. Transient, stress-related paranoid ideation or severe dissociative symptoms."

(American Psychiatric Association, 1994)

Admittedly there are problems with this. These criteria are a mixture of non-verifiable subjective states (such as criterion 7) and more observable behaviour (criterion 8). Furthermore, two patients might be assessed, one meeting criteria 1-5 and the other criteria 5-9. Both would receive a diagnosis of BPD but would only have one criterion in common.

However the nine possible criteria do allow for the possibility of using the number of criteria met as a crude dimensional measure giving a scale of 0-9 in addition to the dichotomous present-absent categorical measure.

4.2.3.4 Summary of position on personality disorder

For the purposes of this study, the working definition of a personality disorder was taken to be that provided by the DSM-IV:

“General Criteria for Personality Disorder

- An enduring pattern that deviates markedly from the person's culture in 2 or more of the following:
 - Cognition (perception & interpretation);
 - Affect (range, intensity, lability, and appropriateness);
 - Interpersonal functioning;
 - Impulse control.
 - The enduring pattern is inflexible and pervasive.
 - The enduring pattern leads to significant distress or impairment in social, occupational, or other areas of functioning.
 - Onset is in adolescence or early adulthood.
 - The pattern is not better accounted for as a consequence of another mental disorder, substance abuse, or another medical condition.”
- (American Psychiatric Association, 1994: p633)

For this study, it was decided to restrict the focus of interest to patients with a diagnosis of Borderline Personality Disorder (BPD), the DSM-IV criteria for which have been given in Section 4.2.3.1 above. The number of criteria met, as well as whether this number crossed the diagnostic threshold, would be recorded as this gave the possibility of a dimensional measure.

Any BPD diagnosis was not to be taken as asserting the existence of an objectively verifiable disease entity, but rather as a useful device for grouping together people with broadly similar, co-occurring patterns of subjective experience and behaviour. These patterns, while on a continuum that would be seen as including the whole population, would be seen to occur relatively rarely and to lie towards the extreme of the continuum.

4.2.4 THE 6PSM IN PERSONALITY DISORDER

The previous two sections have reviewed the literature on the 6-Part Story Method and on personality disorder separately. There are a few publications, none reporting empirical investigations, considering the two subjects together.

The originator of the 6PSM says that: "My assumption is that by telling a projected story based on the elements of fairytale or myth, I may be able to see the way the self projects itself in organised reality in order to meet the world." (Lahad, 1992: p157). While not explicitly mentioning personality disorder, this suggests that Lahad is interested in fundamental, characterological traits of the subject rather than more transient states. Indeed, "the way the self projects itself in organised reality in order to meet the world" could be taken as quite a succinct definition of personality.

Dunn & Parry (1997) have reported on the use of formulations based on the practice of Cognitive Analytic Therapy in the management of people with personality disorder. The service they describe was the first in the UK to routinely use the 6PSM as part of their assessment package, alongside other methods such as self-report questionnaires, interviews and third-party information (Dent-Brown, 1999b). This use of the 6PSM implies that the method was being used to influence (although not absolutely to make) diagnoses of personality disorder – if it were not, then why include it in the assessment battery? This in turn implies a degree of reliance on the 6PSM as both a valid and reliable procedure.

This assumption is at the forefront of a descriptive article (Dent-Brown, 1999a) reporting stories from patients with borderline, schizoid and narcissistic personality disorders. In each case, a story from such a patient is followed by some commentary on the story and how it reveals some aspect of the author's personality. For example, in one story "...there are very clear narcissistic features present. The choice of a main character such as Elvis is characteristic

- other narcissistic patients have chosen characters such as St George, Planet Brain or even God.” (Dent-Brown, 1999a: p13).

The problem here is that these stories have been used to make the initial diagnosis of, in this case, narcissistic personality disorder. Stories from people with the same diagnosis are then being compared and general conclusions are being drawn about all stories that come from people with the diagnosis. There is a dangerous circularity here; if invalid assumptions are being made about what does and does not constitute a “narcissistic feature” then inaccurate diagnoses will be made which will reinforce, rather than challenging, the initial assumptions.

In addition, the stories described are interpreted only by the author, and only one story is elicited from each patient. What if a second interpreter were to inspect the same story; how do we know whether they would reach the same conclusions about narcissistic (or borderline, or schizoid) features? What if a second story were to be elicited from the same patient, with entirely different features present?

These questions highlight the need for an investigation of the reliability and validity of the 6PSM before any further investigation of its clinical use.

4.2.4.1 Potential contribution of the 6PSM to understanding personality disorder

Over the last decade in the increasing attention has been paid to the problem of personality disorder in the UK's National Health Service. A recent Policy

Implementation Guidance produced by the Department of Health notes that

“In many services people with personality disorders are treated at the margins.... Many clinicians and mental health practitioners are reluctant to work with people with personality disorder because they believe they have neither the skills, training, nor resources to provide an adequate service, and because many believe there is nothing the mental health service can offer.” (NIMHE, 2003: p5)

Conversely, NHS patients with this diagnosis report increasing dissatisfaction with existing services, in particular with the way clinical staff communicate with them. In a recent study of service users' opinions, one said:

“Those of us with Personality Disorder can elicit a negative response and a kind of aloofness from professionals, probably because we are a mass of churning emotions and, unintentionally, this is threatening to others.” (Castillo, 2001).

One study of psychiatrist's reactions to case vignettes even suggested that, to quote the title of the article, personality disorder is the label given to “Patients psychiatrists dislike” (Lewis & Appleby, 1988). But how are clinicians to overcome their reluctance to work with this group, when under perceived threat from distressed patients? How can clinicians come to know and understand the

inner experience that drives patients to some of the behaviour they witness and manage in the course of their professional lives?

It has been proposed that many patients with personality disorders normally find it difficult or impossible to communicate their personal stories persuasively. This could be for a number of reasons: "There may be a lot of shame about telling the story. The patient may be anxious... There may be a simple lack of insight.....There may be defence mechanisms of various sorts.... Or there may simply be unfamiliarity with psychological thinking." (Dent-Brown, 1999a: p11). Without an understanding of the patient's narrative, empathy is much harder to achieve and so the position of mutual misunderstanding is reinforced. The suggestion was that the 6PSM could act as a bridge to permit greater understanding and communication between patient and clinician, and to provide clinicians with one of the tools that they reported to be lacking.

4.2.5 CONCEPTS OF RELIABILITY AND VALIDITY

The potential being claimed for the 6PSM by its Israeli originators and those using it in the UK was very large, but untested. Where to make a start on the process of establishing an evidence base? In the author's clinical team, described by Dunn & Parry (1997), the method was used in the context of a psychotherapy assessment. The first test of a newly-developed assessment in psychotherapy is usually a challenge to its reliability and validity – particularly in the case of a psychometric test purporting to make or contribute to a diagnosis. The 6PSM is manifestly NOT a psychometric test, nor is its outcome a dichotomous one such as diagnosis present/absent. Nevertheless it was

decided to adopt the process that might be used in the validation of a psychometric test as the most rigorous evaluation possible. In the context of the 6PSM, what are we to understand by the concepts of reliability and validity?

4.2.5.1 What would be the features of a reliable 6PSM?

Lewis (1999) defines reliability as:

“the degree with which repeated measurements, or measurements taken under identical circumstances, will yield the same results. This definition assumes that the act of measuring does not affect the variable or characteristic of interest. The statistical definition of reliability is related to the lay definition, in that a piece of machinery which is reliable according to the lay definition yields the same behaviour each time it is used. Reliability is a measure of the randomness of the measurement process itself.” (Lewis, 1999: p3)

In the case of the 6PSM the question of reliability refers not so much to the story elicitation method, as to the method of analysis and understanding of the story. The former is well established and uncontroversial, whereas the latter has barely been described. In this context the reliability of these methods might be described as a measure of the consistency of their results. For this study, it was hoped to establish two principal forms of consistency; that between different raters and that between different times. Two key questions to be answered were:

1. What methods of analysing the story maximise the likelihood of two independent raters reaching the same conclusions about a story?
2. What methods maximise the likelihood that two stories from the same person would be analysed or rated in similar ways?

4.2.5.2 What would be the features of a valid 6PSM?

Lewis (1999) proposes that:

“The validity of a measurement can be defined as the degree with which the measured value reflects the characteristic it is intended to measure...

The term ‘valid’ implies that there is some sort of external standard or ‘gold standard’ against which the current measurement is being compared.” (Lewis, 1999: p4)

This poses more of a problem than the question of reliability, as there remains the unanswered question of the status of ‘personality disorder’ and the lack of any gold standard for its measurement. As explained in section 4.2.3.4 above, a pragmatic decision was made to accept the existence of personality disorders as defined by the DSMIV as a heuristic device. Even within this framework there can be no ‘gold standard’ such as a blood test or a chemical assay – therefore what might be termed the ‘silver standard’ of the SCID-II clinical interview was taken as the external yardstick to be used.

There were two forms of validity that were to be investigated, each requiring a different method and hopefully each strengthening the other.

1. First was the question of concurrent validity: even supposing that a reliable method of analysis could be found, do the results of this analysis have any bearing on the personality status of the story authors, as measured by the concurrent instruments used?
2. And second the question of face validity; did the 6PSM produce material that made sense to the authors of the stories themselves?

In a sense these two questions form the two sides of the mixed methodology described in section 5.2.4 following. The answer to the first would be couched in terms of a hypothesis which would be supported or not. The second answer would be couched in qualitative terms and would, hopefully, expand on the meaning of the first answer for the participants themselves.

5.0 PLANNING OF STUDY, METHODS AND ANALYSIS

5.1 OUTLINE RESEARCH QUESTIONS

5.1.1 INTER-RATER RELIABILITY

One obvious gap in knowledge about the 6PSM relates to the reliability of the method. Given a patient producing a story, can two (or more) clinicians agree reliably on what they think that story means, what the patient is communicating, what can be inferred about the author? This presented an immediate problem as there were no adequate systems for reporting on, classifying or otherwise rating a six-part story. The originators of the 6PSM (Lahad, 1992; Lahad & Ayalon, 1993) had proposed one method of rating, relating to the BASIC Ph approach. This proposed that a person's methods of coping with stressful life events could be classified into one of six areas, based on Beliefs, Affect, Social, Imaginative, Cognitive or Physical coping. The suggestion was that instances of the six areas in a story should be tallied, and the area with the greatest number of tallies would be the preferred coping method of the author.

There were two main problems with using the BASIC Ph approach in this study. One was the lack of evidence about inter-rater reliability of BASIC Ph scores; although the originators felt confident that scoring was reliable, no published studies had tested this. Furthermore, reliable classification into BASIC Ph categories would require training not only in the 6PSM tool, but also in the BASIC Ph approach. While the author was experienced in the former, he was not experienced in the latter, which would in any case greatly extend the amount of training needed by any clinician participants to be recruited.

A second problem with BASIC Ph however related to the concept of coping. The approach was developed with normal subjects such as whole school or military populations who had been or might be subject to traumatic incidents. A presumption could be made therefore that most of the participants would have an adequate repertoire of coping skills at their disposal. In the present study it was hoped to investigate the 6PSM in the context of NHS patients with personality disorders; a group of people who might be predicted to have far fewer adaptive coping strategies available.

With this in mind, the first two research questions to emerge were:

- In the context of people with a personality disorder, can a practical rating technique for six-part stories be devised?
- Given this system, can two raters independently rate stories reliably?

5.1.2 TEST-RETEST RELIABILITY

As well as inter-rater reliability, any such rating system would also need adequate test-retest reliability - particularly in the case of a tool that purports to be measuring something about the underlying personality. If the tool is liable to be more sensitive to short-term mental state than underlying personality, its test-retest reliability will be poor.

Throughout his seminal article Frank (1939) makes it clear that he feels it is enduring personality traits, rather than transient states, that influence the form and content of the story. Validity studies of the Rorschach and TAT have most frequently used concurrent measures of personality such as the MMPI for

validation. Others however have questioned this. In a critical article Holmes (1974) describes an experiment where subjects were asked to complete two sets of TAT stories, with identical stimuli, three weeks apart. Very low test-retest reliability was recorded, suggesting to Holmes that either the whole TAT approach was invalid, or that it was tapping transient phenomena and not personality dimensions.

However Holmes's experiment failed to use any concurrent measure of personality; subsequent studies that have done so have demonstrated that it can be a valid indicator of personality style. For example, one study (Shulman, McCarthy, & Ferguson, 1988) demonstrated that participants who scored high on a measure of narcissism following administration of a TAT protocol were reliably identified as showing narcissistic traits by a subsequent blind clinical interview. The recent critical review of projective techniques (Lilienfeld *et al.*, 2000) also acknowledges that with a robust scoring system the TAT can show good validity. For example in a study using the Social Cognition and Object Relations Scale (SCORS) method of rating, it has been shown that patients with BPD score significantly lower on all four SCORS scales than patients without a BPD diagnosis (Westen, Lohr, Silk, Gold, & Kerber, 1990).

It might be asked why the SCORS could not be used as the rating system for the 6PSM, given that both methods produce narrative material and the SCORS had been shown to produce reliable and valid results. The SCORS manual (Westen, Barends, Leigh, Mendel, & Silbert, 1990) was consulted to see whether it would serve, but it became clear that the method required not only expert training but also a degree of familiarity with Object Relations theory.

Neither could be assumed among the body of raters to be assembled for this study, although it was encouraging to find that story material could be rated in this way and produce statistically significant results when applied to personality disorder.

After these considerations, a third research question was now clearer:

- Are the ratings of stories from the same authors reasonably stable over time, with the possible implication that they are tapping relatively stable characteristics rather than more transient states?

5.1.3 RATER BIAS

It has already been mentioned that when a clinician peruses a story looking for the author's projections, they may instead find their own. It was unknown to what extent this phenomenon might contaminate any ratings of a six-part story. The establishment of inter-rater reliability might suggest that at least some part of the rating was not dependent on the characteristics of the individual rater, but it was hoped to investigate this possibility in more detail. The research question was:

- Do the ratings given by raters to others' stories bear any relation to the personal characteristics of the rater themselves?

5.1.4 CONCURRENT VALIDITY

Even if inter-rater and test-retest reliability could be established through some new rating scheme, would these reliable ratings allow any inferences to be

made about the stories' authors? For example, maybe raters could agree reliably on some rating of the sociability of the main character, and perhaps this would be stable from one story to the next. But would this rating enable us to say anything about the sociability of the author him or herself? This led to the next research question:

- Do the results from the rating method/s devised for the 6PSM bear any relation to other data about the stories' authors - for example personality disorder diagnosis, general mental health or gender.

5.1.5 PARTICIPANTS' EXPERIENCE OF THE 6PSM

One of the clinical reasons behind the use of the 6PSM in the author's own psychotherapy service was the assumption that this rather unusual method (by comparison with more common interview approaches) might yield new and useful information. In particular it was hoped that the 6PSM might enable patients with a personality disorder to communicate something about themselves metaphorically. This extended metaphor might permit a better communication of the patient's underlying personality process because a fictional story would not have the burden of autobiographical content that usually makes up patient narrative in the clinical setting. It was also believed from clinical experience that patients found the 6PSM an acceptable, albeit rather unfamiliar, method of assessment.

All of these assumptions seemed to be borne out in clinical practice, but patients undertaking the 6PSM had not been asked systematically about their

experiences of the method. It was decided to try to address two further research questions to do with patients' subjective experiences:

- How acceptable is the 6PSM to patients as a method for general clinical use?
- To what extent does the 6PSM produce material that has face validity for the authors themselves?

In addition it was hoped that data gathered to answer these questions might allow a third to be answered:

- Do the subjective accounts of authors of a six-part story permit the development of a theory about the mode of action of the 6PSM?

5.2 METHODOLOGICAL AND EPISTEMOLOGICAL CONSIDERATIONS

The process of devising the research questions did not, of course, come to a conclusion without some ideas about appropriate research methods arising. As each question became clearer, possible ways of answering it were considered. As far as possible however the research questions were allowed to dictate the methodology of the overall study, rather than selecting a methodology and then considering what questions could be answered by using it. It is for that reason that this section follows section 5.1 on developing the research questions, rather than preceding it.

5.2.1 PREFERRED METHODOLOGIES IN DRAMATHERAPY RESEARCH

Unsurprisingly, writers on methodological matters tend to agree that the approach outlined above is the right one. Even in a field like dramatherapy with a very recent and narrow evidence base, it is apparently accepted that no one methodology can on its own successfully answer all questions:

“Different research paradigms are not simply different versions of the same activity: they regard life and the world in different ways. This, however, is precisely why they are so useful. Taken together they provide us with a range of ways of increasing our understanding...”

(Grainger, 2002: p9)

“...if we are going to play in the field of research we need to understand many research methods, appreciating the limits and strengths of each, so that we are able to make a fit between the models selected and the particular needs of the paradigm under investigation.” (Barham, 2003: p6)

Unfortunately it does not seem to be possible to consistently hold this catholic attitude to the selection of methodology. The two authors cited immediately above are well known to UK dramatherapists, but elsewhere Grainger comments on a piece of his own quantitative outcome research, saying:

“We settled for clarity, even if it showed us what we already knew and did not really want telling.... The main justification for carrying out quantitative research in dramatherapy, in our opinion, is to convince the sceptical that it can produce a measurable degree of healing change in psychologically disturbed and vulnerable people.” (Grainger, 1999: p20)

And in describing the research activities of his academic department, Barham says:

“...we have a mainly qualitative approach as we believe passionately in using methodologies that are congruent with the way that as artists and therapists we both see the world and engage with it.” (Barham, 2003: p4)

Both authors appear to endorse using a wide range of methodologies in principle, while in practice privileging qualitative methods. Quantitative methods are seen as largely unnecessary for dramatherapists, and incongruent with the way dramatherapists experience the world. Another well known author on research in the arts therapies suggests a third route in addition to qualitative and quantitative methodologies; that of art-based research (McNiff, 1998). He proposes that research in the arts therapies can best be conducted through the medium of the art itself:

“I tend to achieve insights in an experiential way, through... the 'genuine truth' of aesthetic understanding. When I ascertain something in this way, I do not generally need quantitative verification to affirm my belief.”

(McNiff, 1998: p148)

McNiff notes approvingly that most art therapy research in the past has taken the form of single case studies, and feels this primacy should be reinforced and the case study used as the default form of art therapy enquiry (McNiff, 1998: p141). Although McNiff writes as an art therapist, his call for the primacy of art-based research has been welcomed by some dramatherapists (Grainger, 1999pp.128-131).

5.2.2 QUALITATIVE OR QUANTITATIVE APPROACHES?

In this study it was decided to adhere to the principle of fitness for purpose in selecting the methodology and techniques to be used; those would be chosen which were most likely to provide a useful and convincing answer to the questions posed. Support for this approach came in a source that was published once the study was well under way. Salmon (2003) is discussing research in psychology, not dramatherapy, but he describes a similar tension when he writes that:

“Articles...have passionately argued whether it [qualitative research] will save the discipline from the perils of positivism, or herald its banishment from the community of sciences into subjectivity and obfuscation.”

(Salmon, 2003: p24)

Salmon describes Feyerabend's (1975; 1978) notion of the 'anarchist scientist', whose epistemology "... states not that rules are unnecessary but that all methodologies have limits..." (Salmon, 2003: p25). He goes on to suggest that "... the fundamental axioms of quantitative and qualitative enquiry are arbitrary: whether to be quantitative or qualitative in any specific study should be decided by 'fit' with the phenomenon being studied." (p.25).

Maxwell (1996) proposes that:

“Quantitative approaches are powerful ways of determining *whether* a particular result was causally related to one or another variable, and to what *extent* these are related. However qualitative research is often better at showing *how* this occurred.” (p.59, italics in original).

This description potentially puts the two approaches not into opposition, but into roles complementary to one another. Applying this very simply to the research questions outlined in section 5.1, it was decided to adopt a quantitative approach to the question of *whether* the 6PSM reliably produced meaningful material, while a qualitative approach would be taken to investigating both *how* it did this, and *how* the method was received by research participants.

5.2.3 EPISTEMOLOGICAL POSITION

Adopting a pragmatic, rather than a principled, approach to the choice of methodologies does beg the question of the philosophical positions underlying the study. Was the study to adopt the positivist position that 'the truth is out there', or the interpretivist position that the data are constructed by, rather than revealed by, the researcher?

A metaphor was suggested by the work of Winnicott, in particular his descriptions of transitional objects. For example, in Playing and Reality he says:

"Of the transitional object it can be said that it is a matter of agreement between us and the baby that we will never ask the question: 'Did you conceive of this or was it presented to you from without?' The important point is that no decision on this point is expected. The question is not to be formulated." (Winnicott, 1971: p12)

It is precisely from this refusal to answer the question, Winnicott argues, that creativity springs. In the context of this study, it was hoped to leave a creative

'research space' (analogous to Winnicott's 'play space') between the positivist and interpretivist paradigms. The measure of the results of the research would not be the purity of their origins, but their degree of interest and usefulness. This concept of playfulness was echoed by Salmon:

"...the anarchist scientist is playful rather than precious with methodology. Methods are chosen simply because they interest the researcher or because the researcher values their products." (Salmon, 2003: p25).

That being the case, existing qualitative and quantitative studies were next consulted to see whether their methods were interesting or their results valuable to the present research questions.

5.2.4 RATIONALE FOR SELECTION OF A MIXED METHODOLOGY

Spradley (1979) proposes the following contrasts between qualitative and quantitative approaches:

Table 5.1: Contrasts between quantitative and qualitative approaches

Research with Subjects (Quantitative)	Research with Informants (Qualitative)
1. What do I know about a problem that will allow me to formulate and test a hypothesis?	1. What do my informants know about their culture that I can discover?
2. What concepts can I use to test this hypothesis?	2. What concepts do my informants use to classify their experiences?
3. How can I operationally define these concepts?	3. How do my informants define these concepts?
4. What scientific theory can explain the data?	4. What folk theory do my informants use to explain their experience?
5. How can I interpret the results and report them in the language of my colleagues?	5. How can I translate the cultural knowledge of my informants into a cultural description my colleagues will understand?

Some authors (Lincoln & Guba, 1985) argue that these two approaches are fundamentally incompatible and cannot be jointly adopted. The reason given is that the questions above do not simply require different methods, but that they proceed from fundamentally different views about the nature of knowledge. Lincoln & Guba propose that all knowledge is socially constructed and that variables are complex, interwoven and not reducible to numbers or generalisable beyond their context. Thus all of the questions on the left of the table above are in their view mistaken and any answers derived will be misleading.

Other authors such as Cresswell (1994) give primacy to pragmatic, rather than philosophical considerations. They argue that the two approaches can be amalgamated because the methods of investigation can in fact be similar, and that there is a continuum, rather than a dichotomy, between quantitative and qualitative methods.

It was noted in section 4.2.5.2 that there were two questions of validity in this study; one couched in quantitative and the other in qualitative terms. Neither was given priority in the study and each was regarded as of equal and complementary importance. Because of the desire NOT to privilege one form of investigation over another, Cresswell's amalgamated, continuum-based approach was preferred.

There were also the dual roles of the participants. In terms of the table 5.1 above, participants were to be treated both as informants and subjects, not exclusively one or the other. Participants were to be seen as expert informants with their own theories, as well as subjects to provide data for external analysis by the investigator.

For these reasons a mixed-model, quantitative/qualitative approach was adopted.

5.3 LESSONS FROM PUBLISHED STUDIES

5.3.1 QUANTITATIVE STUDIES

The majority of the research questions set out in section 5.1 were very similar to questions about the psychometric properties of projective and other assessment tools. Despite the initial resistance to statistical methods of establishing the reliability and validity of projective approaches, there is now a large body of literature describing such actuarial methods. What could be learned from these reliability and validity studies of other projective approaches?

5.3.1.1 Projective storymaking approaches to personality

The 6PSM is not the only projective tool based on storymaking to emerge in recent years. The original TAT approach has spawned offshoots such as the TEMAS Tell-Me-A-Story test (Costantino & Malgady, 1999). The TEMAS was specifically developed to counter the criticism that the TAT was very specific to a white, English-speaking US context. It uses the same principle of having picture cards as a stimulus, but differs from the TAT in using multicultural themes featuring Hispanic and black as well as white protagonists. It has been validated for use with children of 5-14 years and it has been suggested that when used alongside objective (teacher-rated) measures of children's behaviour each method assesses different aspects of functioning and that the two measures are therefore complementary (Flanagan, 2000).

Another measure designed for use with children is the Fairy Tale Test (Coulacoglou, 1995; Coulacoglou & Kine, 1995). The FTT also relies on cards presented as stimuli. In this case the cards are in seven sets of three, each set of three showing different versions of the same character, for example Snow White, a giant or a witch. Participants are asked to compare and contrast the images with questions such as "Which of the three giants is the most wicked? Why?" Normative data for the FTT have been collected from a group of 800 8-12 year old children from schools in Greece.

Perhaps because story-based approaches are seen as more suitable for children, they have been applied to adults less often. One method has been reported for assessing the attachment style of adults (West, Pettem, Sheldon-Keller, Rose, & Cawthorpe, 1995) which again uses drawings - in this case three line drawings of attachment situations.

One possible problem with all three of these approaches, as well as the original TAT, is the use of a picture as a stimulus. While considerable care is usually taken in the construction and selection of stimuli, the doubt must remain that the participant's response may say more about the stimulus than their own personality, attachment style or level of function. Another problem specific to the TAT arises from the number of TAT cards in use. Because there is little consistency in which cards are selected for any given study it has been argued (Lilienfeld *et al.*, 2000) that the generalisability of TAT scoring schemes is poor. This second criticism does not however apply to the three storymaking methods described above.

The 6PSM however avoids this pitfall of the TAT and other picture-stimulus systems because there is no visual prompt. Instead the 6PSM uses a series of verbal prompts, such as:

“Now in your first space I want you to draw a picture or sketch of the main character of your story. This need not be human – it could be an animal or a fantasy character or anything that you could imagine might be able to talk and think and feel.”

5.3.1.2 Validity and reliability testing of existing storymaking assessments

The Tell-Me-A-Story (TEMAS) test and the Fairy Tale Test (FTT) have both been the subject of validity and reliability testing (Coulacoglou & Kine, 1995; Flanagan, 2000; Flanagan & di Giuseppe, 1999). What can be learned from these studies to inform a similar testing of the 6PSM?

Flanagan & di Giuseppe (1999) report a sample of 642 schoolchildren from whom normative data for the TEMAS were developed. Children's stories were scored by trained raters, and the sub-scales had an internal consistency ranging from .62 to .83, which is satisfactory. However the inter-rater reliability of the scores making up these sub-scales was itself rated as moderate at best, with some items having an inter-rater correlation of scores as poor as .33. This was statistically significant, given the large sample, but is of arguable clinical significance in any individual case. Test-retest reliability over an 18-week period was described as poor to moderate, although no values were given in the text of the article. It was said that “...this may be attributed to possible inconsistency on the part of the subjects and the use of different raters at pre-test and post-test”

(Flanagan & di Guiseppe, 1999: p25). It is somewhat odd to identify the inconsistency as lying with the participants, rather than the raters or the instrument, but there is a further weakness here.

The TEMAS scoring system passes the first test of internal consistency well enough, but if different scorers are used to collect the pre- and post-test data then reasonable inter-rater reliability is a pre-requisite. Without this, it is not surprising that test-retest reliability cannot be demonstrated and without this any assertions about validity must be questionable.

Flanagan & di Giuseppe do nevertheless report on the validity of the TEMAS, reporting that 86-89% of cases were correctly classified as clinical or non-clinical. They do concede that the definitions of clinical and normal are problematic, given that no single gold standard was used across all participants to give a definitive clinical diagnosis in advance of the TEMAS being applied. In a separate article Flanagan (2000) reports the correlation between results from the TEMAS and another assessment of children's behaviour, the Behaviour Assessment System for Children (BASC). She indicates that the correlations, while statistically significant, are low (only one of the nine reported correlations exceeds .35) and the total number of non-significant correlations is not reported. This apparently low concurrent validity is rationalised by suggesting that the BASC and the TEMAS are measuring different things and may therefore be complementary and should be used together

The Fairy Tale Test (FTT) also has normative data from a large group of children (n=800) (Coulacoglou & Kine, 1995). There are no data reported here

on the internal consistency of the scales that were constructed, but inter-rater reliability was tested on a sub-sample of protocols (n=49). This appeared to be good, statistically significant on all 19 sub-scales and ranging from .47 to .96 with most scales above .80. Test-retest reliability was also assessed on a sub-sample (n=52) and was described as '...moderate (.4-.7) for the majority of variables.' (Coulacoglou & Kine, 1995: p19).

Validity was assessed against parent and teacher ratings of the children involved, and all children were rated using the same rating scheme. This gives a better, independent measure of clinical status than was available for the TEMAS. A second measure of validity was to group the stories into those described as 'disturbed' and others, and to compare the scores given to the two groups. However the initial division of stories into two groups was made purely by the researcher, who then went on to score the stories in more detail. The fact that the scores of stories from the two groups differ is not therefore so impressive as if the division of stories and their rating had been carried out independently.

The studies of the TEMAS and the FTT have admirably large normative groups, numbers which it was impossible to match given the resources available to the present study. It was encouraging that some progress towards acceptable levels of validity and reliability was made, but there also seemed to be some flaws in the studies, such as:

- Internal consistency of scales either not tested or not reported (FTT)
- Low inter-rater reliability possibly compromising test-retest reliability (TEMAS)

- Scales with low test-retest reliability nevertheless being used for validity testing (TEMAS)
- No gold standard for testing concurrent validity (TEMAS)
- Division into groups and scoring of groups not carried out by independent raters (FTT)

As a result it was decided to approach a validity and reliability study of the 6PSM in the following way, to take the best elements of the TEMAS and FTT studies and learn from their weaknesses:

- Individual items to be tested for inter-rater reliability as a first step
- Only those items passing this test to be assembled into scales maximising internal consistency
- Scales to be tested for inter-rater reliability and test-retest reliability
- Only those scales passing both these tests to be used for validity testing
- Gold standard tests (as far as they exist) to be used concurrently on all patient participants
- Application of concurrent tests and rating of story material to be carried out independently by separate individuals

Thus it was expected that even if the total numbers of stories to be rated were smaller than in the TEMAS and FTT studies, the rest of this 6PSM study would be methodologically sound.

5.3.2 QUALITATIVE STUDIES

The research question that seemed most open to investigation using a qualitative approach was the last one generated:

- Do the subjective accounts of authors of a six-part story permit the development of a theory about the mode of action of the 6PSM?

Miles & Huberman (1994) refer to theory-building studies as belonging to the realm of social anthropology. They say that “researchers in life history, grounded theory, ecological psychology, [and] narrative studies ... often take this general line.” (Miles & Huberman, 1994: p8). This led to looking more closely at Grounded Theory first developed by Glaser & Strauss (1967) which is characterised by attention paid to a number of factors:

“... the relevance of theory, grounded in data, to the development of a discipline... ; the belief that persons are actors who take an active role in responding to problematic situations; the realisation that persons act on the basis of meaning [and] the understanding that meaning is defined and redefined through interaction...” (Strauss & Corbin, 1998: p9)

These assertions were congruent with the way the 6PSM had been used in the author’s clinical team, and it was decided to use the methodology of Grounded Theory to inform the qualitative arm of the study.

No Grounded Theory studies of projective approaches could be found, but there were two articles describing a Grounded Theory study of transitional objects (Arthern & Madill, 1999, 2002) that were closely analogous to this study. The aim was to use interviews with therapists and their patients about their

experience of using a transitional object in psychotherapy; the transcripts of the interviews would be analysed using the Grounded Theory method of coding and constant comparison to develop a theory to explain the mode of action of transitional objects.

Many other accounts of investigations of story-based and narrative techniques were also found, but these were not so directly helpful. For example, Abma (1998) produced a formal qualitative investigation using patient narratives about their hospital experience to reveal features about the patient-staff relationships. However here the story was the tool of the enquiry, not the subject of it.

McArdle & Byrt (2001) report on the therapeutic value of writing stories, but in a literature review of clinical descriptions rather than a systematic review of research studies.

Lacking any other studies, the Arthern & Madill study was taken as a starting point for thinking about designing this piece of research. Arthern & Madill recruited a group of six psychotherapy patients who had all used transitional objects (TOs) during their therapy, and then interviewed both the patients and their therapists about their experiences. This produced 12 lengthy in-depth interviews which were then analysed using the constant comparison method to produce a range of concepts, categories and a final central category. Constant comparison is a central technique in Grounded Theory development, and involves the formation of tentative hypotheses (categories) which are constantly checked against the data as the analysis progresses in order to check their validity and to modify them as new data are analysed, As interviews progressed, those concepts that emerged in early interviews could be followed

up, alongside newly emerging concepts. The end result combined a rich account of the experience of using TOs with a coherent theory for their action in therapy. It was hoped to use elements of this method in developing a Grounded Theory of the mode of action of the 6PSM.

5.4 OUTLINE PLAN OF RESEARCH PROGRAMME

As the majority of the research questions were to be answered by quantitative means, these methods were addressed first. It was decided to collect, by tape recording and transcription, as many stories as possible from patients of the adult CMHTs in the author's NHS Trust. Some would be gathered from patients with a likely diagnosis of Borderline Personality Disorder (BPD) while others would come from more mainstream CMHT patients. This would allow for similarities within the groups and differences between the groups (if any) to be explored.

5.4.1 PLANNING OF PILOT STUDY

The key problem here was the conversion of tapes, transcripts and images produced in the 6PSM process into a form that would permit of quantification. A pilot study was planned using two existing stories from the author's clinical work and a panel of volunteer raters who had previously attended training courses in the 6PSM run by the author. A variety of rating techniques would be tested out, including asking raters to categorise stories or their authors, presenting raters with descriptive statements of the story with which they could agree or disagree on a Likert scale, asking raters to make ratings of the inferred mental health of the story's author and so on.

If a robust scoring method of some sort could be devised, it was then hoped to be able to answer the question about rater bias raised in section 5.1.3. The rating would be carried out by the same clinicians who administered the 6PSM

to patient participants. The clinicians themselves, as described in the next section, would already have completed their own 6-part story, so it might be possible to compare the ratings they made of others' stories with the ratings others made of theirs.

5.4.2 RECRUITMENT OF PARTICIPANTS

While the piloting was going on, the process of recruiting CMHT patients was started. When this was considered in more detail, the question of who should elicit the story from the patient arose. If clinicians recruited the patients, but the 6PSM session was undertaken by the researcher, there would be no problem with a consistency of approach, instructions etc.; all participants would receive a very similar experience. However this would be unlike the use of the 6PSM in clinical practice, where the technique would be used in the context of an ongoing relationship. There was also a potential ethical problem of what to do if therapeutically important material was elicited by the researcher who had no clinical relationship with the patient.

This resulted in the decision to ask clinicians to elicit 6-part stories from their patients, using a script to be developed and asking them to stick exactly to the script for the portion of the session that was to be taped and analysed in the research. This required that, even with a script, clinicians needed training in the administration of the 6PSM, which presented the opportunity of taping 6-part stories produced by clinicians during their training and adding these to the transcripts to be analysed.

5.4.3 DATA TO BE COLLECTED

Each clinician would be asked to tell a story during their training and then recruit two patients from their caseload - one of whom met outline criteria for a possible BPD diagnosis and one of whom did not. The clinician would then record a 6-part story from each patient, and record a second story from the two patients after a one month gap. This would allow comparisons to be drawn between stories from the three groups: clinicians, patients with possible BPD and mainstream patients. Recording two stories from each patient would allow test-retest reliability of the 6PSM and the developing rating scheme to be assessed.

In between the two story recording sessions a single session to gather some concurrent measures was planned. This would enable some triangulating assessments of the patients to be made, against which their possible BPD diagnosis could be checked and their 6PSM data compared.

5.4.4 CONCURRENT MEASURES TO BE USED

Given that patient participants would already be giving up two hours of their time for story recording, it was felt necessary to complete all the concurrent measures in a single session. The measures were therefore selected with a view to it being possible to complete them all in a maximum period of two hours.

5.4.4.1 SCID-II

A key measure was to be some way of assessing the nature, degree and comorbidity of any personality disorder. The Millon Clinical Multi-Axial Inventory (MCMI) (Millon, Davies, & Millon, 1997) was considered for this. The MCMI is a 175-item self-report questionnaire closely aligned to the DSM-IV. There are scales for each of the DSM-IV personality disorders and normative data available. In addition, the MCMI had been used in the author's department and had been used on many patients with a BPD diagnosis.

However there were some drawbacks. The instrument takes a minimum of 25 minutes to complete, and requires a good level of literacy of anyone completing it (although the questions can be presented verbally if necessary.) Experience in the department had shown that certain scales did not seem to be as sensitive as others; for example the scale for Narcissistic Personality Disorder was very rarely scored above the clinical threshold, even by patients with a clear narcissistic disturbance. This may have been because the instrument was developed in the USA and there may be some social or cultural assumptions in the questions that make it less suitable for UK use. Alternatively, it may have been because self-report is not the most reliable way of assessing personality disorder, either because of lack of insight on the respondent's part or because of lack of candour.

It has been noted that self-report measures are less reliable than structured interviews in this situation (Jackson, Gazis, Rudd, & Edwards, 1991). Two such interviews were considered; the International Personality Disorder Examination,

IPDE (World Health Organisation, 1995) and the Structured Clinical Interview for DSM-IV axis II personality disorders, SCID-II (First, Gibbon, Spitzer, Williams, & Benjamin, 1997). The latter has versions for both the DSM-IV and ICD-10, but is the longer of the two. The SCID-II also has an advantage in having a screening questionnaire to be completed by the participant. This speeds up the interview because questions need only be asked about the problem areas the participant has indicated; it has been designed to be over sensitive so that problems reported as absent can reasonably be relied on to need no further probing. This combination of initial self-report followed by clinician assessment has been shown to be an efficient, sensitive and specific approach (Jacobsberg, Perry, & Frances, 1995).

The SCID-II provides a categorical, present/absent diagnosis for each of the DSM-IV personality disorders, on the basis of whether or not the threshold number of criteria has been reached. For example, there are nine criteria for a diagnosis of BPD, and a diagnosis is made if any five are met. In addition to the binary present/absent result, it is therefore also possible to envisage a 10-point (0-9) scale of the number of criteria met. The global number of all SCID-II criteria met throughout the whole interview may also act as a crude measure of overall personality disturbance.

Rogers (2001) reports on 11 studies of the reliability of the SCID-II. Inter-rater reliability estimates range from .67 to .91, while test-retest reliability ranges from .49 to .62. These results suggest that while raters can agree with one another on diagnoses at any point in time, these diagnoses tend to be less stable over time. Test-retest periods varied from 2 to 5 weeks in the studies surveyed.

Torgersen, Skre, Onstad, Edvardsen, & Kringlen (1993) conducted a large scale study of the SCID-II involving 445 participants, involving a factor analysis of the criteria reported, which strongly supported the typology of the SCID-II. Two thirds of the criteria examined loaded first onto the predicted personality disorders, suggesting good construct validity.

Convergent validity of the SCID-II is more problematic to determine, as there is no gold standard against which it can be measured. Even if it measures poorly against some other instrument, it cannot be said with certainty which of the two is performing better! Rogers (2001) reports studies of the SCID-II and MCMI with correlations on the dimensional scales between .30 and .46. Jackson *et al.* (1991) reach the same finding, with the interesting exception of the borderline category, and conclude that it is the SCID-II that is the more reliable instrument.

5.4.4.2 IIP-32

The Inventory of Interpersonal Problems (IIP) (Horowitz, Rosenberg, Baer, Ureno, & Villasenor, 1988) was devised to tap the difficulties people experience in interpersonal relationships. It has been increasingly adopted by psychotherapy researchers as a measure of outcome, and as very many of the DSM-IV Axis II criteria relate to interpersonal difficulties it was felt that it would be an appropriate concurrent measure.

The IIP has been shown to be sensitive to the presence of personality disorder (Scarpa *et al.*, 1999) but at 127 items was too long to be completed in the

timescale of a session which would also include administration of the SCID-II and another instrument. However the IIP-32 (Barkham, Hardy, & Startup, 1996) is a 32-item version that is much quicker to complete but sacrifices little when compared to the longer version.

The IIP-32 produces scores eight paired scales (eg two paired scales are labelled "too aggressive" versus "hard to be assertive") and a summed global score. Some of these scales would appear to be highly relevant to BPD, for example the "too caring" and "too dependent" scales. Indeed, the original IIP has been used as a concurrent measure in other studies of patients with borderline difficulties (Leichsenring, Kunst, & Hoyer, 2003).

Barkham *et al.* (1996) report that for a clinical population, age and sex differences have little impact on IIP-32 scores. This was not the case for a normative GP practice sample, but as the current study involved clinical samples this was not a problem. Internal consistency was good; the alpha coefficient varied between .64 and .84 for the sub-scales and .87 for the global score. Test-retest stability was also good, .56-.81 for the subscales and .70 for the global score.

The IIP-32 has been used in conjunction with measures of borderline personality structure and sexual aggression (Leichsenring *et al.*, 2003; Moriarty, Stough, Tidmarsh, Eger, & Dennison, 2001), with significant (though different) associations in both cases. A degree of convergent validity was therefore felt to exist, and it was decided that the IIP-32 was a suitable instrument for this study.

5.4.4.3 CORE-OM

The Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-OM) was developed as an outcome measure for use in psychotherapy and counselling settings (CORE System Group, 1999). Its development was prompted by the growing need in the UK for a method of comparing individual patients and whole patient groups in a way which could be meaningfully compared from one study to another, for example to compare outcomes across the patient populations of two or more services in a benchmarking study.

The CORE-OM is a 34-item self-report questionnaire with items grouped into four domains: Wellbeing (quality of life statements), Functioning (ability to perform daily tasks satisfactorily), Problems (symptom statements) and Risk (to self or others.) In this study it was believed that the Risk sub-scale in particular would be scored more highly by participants with a diagnosis of BPD, and that there might be other patterns detected in participants' responses in other domains.

The internal consistency for the CORE-OM and its subscales was tested out on non-clinical ($n = 1009$) and a clinical ($n = 713$) samples. In both cases internal consistency was very similar, ranging from .75-.90 for the sub-scales and reaching .94 for the global score. Test-retest stability was measured on a small non-clinical sample ($n = 43$) and ranged from .64-.88 for the sub-scales and .90 for the global scale.

Convergent validity was tested against several well known instruments such as the Beck Depression Inventory (BDI), Symptom Check List (SCL-90), General Health Questionnaire (GHQ-28) and IIP-32. The strongest relationships were between the CORE-OM sub-scales and certain of the other instruments. For example, the CORE Risk subscale correlated moderately with the GHQ severe depression subscale ($r = .69$) while the CORE Function sub-scale correlated best with the IIP-32 ($r = .65$). This suggests that the CORE is indeed tapping a common core of problematic states with convergent validity against longer, more specific measures.

5.4.5 SUMMARY PLAN FOR QUANTITATIVE ARM OF STUDY

The outline plan for the quantitative arm of the study was now clearer. Up to 30 CMHT clinicians would be recruited, each of whom would record a 6-part story as part of their training. Each clinician would recruit two patients from their caseload, one meeting some outline criteria for a possible diagnosis of BPD and the other a more mainstream patient not meeting those criteria.

Each patient would record two stories with their clinician, one month apart.

Between the two story sessions the researcher would meet each patient once to conduct a SCID-II interview and ask the patient to complete a CORE-OM and IIP-32 self-report questionnaire.

A pilot study would hopefully establish a means of quantifying the stories in a meaningful way. Following this the stories would be transcribed and then rated

by a panel of raters drawn from the same clinicians who had undertaken the story sessions.

The precise details of how the concurrent measures were to be related to the story material could not be decided at this stage, because the pilot had not yet been completed. However in outline, if categorical story data resulted from the pilot these could be compared using techniques such as *t*-tests, chi-squared tests, percentage agreement between raters, Kappa and so on. If ordinal or continuous data resulted from the pilot then all the above would be possible alongside methods such as Pearson's or Spearman's correlation.

5.4.6 QUALITATIVE ARM OF STUDY

The majority of the research questions decided on seemed open to a quantitative approach, so to some extent the qualitative arm of the study was subordinated to this. It was hoped to gather data for a Grounded Theory analysis in two ways. First, as part of the script to be read to patients by clinicians, there would be some questions about the storyteller's subjective experience of the story. This would provide the same stimulus questions to every storyteller, with a small advantage in consistency but a major loss of flexibility. In addition to this therefore it was planned to organise some small focus groups of patients to ask them to discuss the experience of the 6PSM and begin to hypothesise about its mode of action. With clinicians this was easier to organise, as they would be trained in twos and threes and a short discussion like this could form the end of every training session.

The method for carrying out a Grounded Theory analysis is set out in detail by Strauss & Corbin (1998), and it was decided to apply this as completely as possible. Any other decisions about analysis would be premature at this stage, as these would have to be made on the basis of the emerging data.

6.0: PILOT STUDY

6.1 PLACE OF THE PILOT WITHIN THE STUDY AS A WHOLE

Before the main study was attempted, a pilot study was undertaken to test out some possibilities and inform the design of the main study. The plan for the study as a whole was as follows:

Preparation and planning:

1. Gather pilot data to establish feasibility.
2. Plan main project and gain ethical approval.
3. Establish protocol for administering 6PSM to subjects.

Data gathering:

4. Recruit clinician and patient subjects. Train clinicians in protocol and have them record 6-Part Stories from clinicians and patients.
5. Gather concurrent clinical material from patients (questionnaires and interview.)

Data analysis:

6. Compare different methods of rating a 6-Part Story.
7. Identify those methods with the best inter-rater reliability and formalise 6PSM rating instrument/s using these methods.
8. Establish test-retest stability.
9. Investigate possibility of rater bias.
10. Validate 6PSM rating instrument by comparison with concurrent clinical material.

This chapter describes the method and results of the first of these 10 stages.

6.2 AIM OF PILOT STUDY

While the main study was being designed and ethical approval obtained, a pilot study was undertaken to begin to identify which methods of rating were potentially useful and offered some possibility of inter-rater reliability. The basic premise was that raters should be exposed to the picture, tape recording and transcribed text of a 6-part story. The raters would be blind to any other information about the author of the story, their diagnostic status, mental health or any demographic data. Raters would then be asked to give their responses to or ratings of the story in a number of ways. A range of different response types was planned (5 point Likert scale, selection from fixed range of categories, free text response) to see what kinds of information could be derived from each. This meant a potentially lengthy instrument, much of which would eventually be discarded. For this reason it was decided to ask a relatively large group of raters to rate just two stories.

6.3 DEVELOPMENT AND METHOD OF PILOTING

6.3.1 SELECTION OF STORIES FOR PILOTING

Two patients were identified from the researcher's clinical practice. They were approached because each had different presenting problems and had already completed a 6PSM as part of their normal assessment process. Both patients had been referred for assessment to the tertiary NHS personality disorder service described by Dunn & Parry (1997). One patient fulfilled the DSM-IV Axis II criteria for Borderline Personality Disorder (BPD), while the other did not. Her

problems were better described by an Axis I diagnosis of Obsessive Compulsive Disorder. It was hoped that the difference in clinical presentation would imply a degree of difference in the stories produced by each patient. The assessments were made by clinical interviews without the use of more formal tools such as the SCID-II developed by First, Gibbon, Spitzer, Williams & Benjamin (1997) or any self-report tests. Both patients were female, in their thirties and twenties respectively.

The session when each patient produced and told her 6-part story had already tape recorded and transcribed (including the researcher's questions.) The recording and transcription were normal parts of the assessment, and not undertaken specially for the research. Each patient was approached and asked if this material could also be used in the research study. They were given written information about what would be done with their material, and signed a consent form agreeing to this.

The material assembled for each rater comprised:

- Copies of the pictures drawn by the two patients during the storymaking process.
- Copies of the original tape recordings of the stories and discussions about them between patients and researcher (22-23 minutes long).
- Transcripts of the tapes (approx 4,000 words each).
- A 12-page response form (described below) for each story.

An abbreviated form of the stories produced by each patient is given overleaf.

Story 1 from patient with Axis II (BPD) diagnosis.

Once upon a time there was a starfish, the sea had brought it in and it was stranded on the beach and it needed to be put back in the water. So Somebody came along and picked it up and put it back in the water so that it'd have a chance of surviving and... but... the danger when the tide comes in there's rocks and there's water round them and the starfish get stuck in them, and when the heat dries the water out the starfish'd die. So you have to go round all the rocks and look at them to make sure there's no starfish that need help.

What would help? A person. Rather the tide coming back in, that's why it comes up a bit more there. The tide coming back in to help it to carry the starfish away, or somebody walking who'll pick it up and put it in the sea. Then the person can stand and watch the starfish... wiggle away and swim and then when you know it's safe and got back to all its family because you don't see it at all in the sea. It's a daft story.

Story 2 from patient with Axis I (OCD) diagnosis

Er... right. Once upon a time there was a very large, smiley sun, whose job - oh I'm onto the next page already - whose job was to shine on all the corn in the field. All right? (Laughs) It was a very big field of corn and it used to blow in the wind. It was quite a difficult job, but it was the sun's job and it had to do it every day. (Laughs) You want me to carry on, yeah?

There was this horrible nasty black cloud that used to come along and rain on the corn and stopped it from growing. And the weeds that used to climb up and strangle the corn, and the insects that used to eat away at the corn. But the sun used to come out every day, well try and come out every day, and shine down on the corn to make it grow. Also helping the corn grow was the fertiliser the farmer used to put down and he used to come along and check on the corn every day to make sure it was growing properly. Right? Am I doing all right? Or do you want some more detail?

He always used to worry that maybe the crop wouldn't grow very well, but in the end the crop always managed to grow quite well and succeed and become healthy. And the end of the story was that because the farmer's corn used to grow really well there was always food on the table to eat. That was it really.

6.3.2 PILOT RESPONSE FORM

6.3.2.1 Quantitative Data

It was intended that a major part of the response form would comprise statements that could be responded to by raters on a five-point Likert scale (strongly agree – strongly disagree.) It was hoped that items displaying adequate inter-rater reliability could then be subjected to factor analysis and assembled into scales for the main study. Potential items for the response form

were developed in the following ways. Two other members of the personality disorder service were interviewed by the researcher. The service has used the 6-Part Story Method (6PSM) since 1996 and both interviewees were very familiar with its use. They were asked to describe their experience of using the 6PSM with patients. In particular, they were asked what features they would expect to find in a story produced by a patient with a borderline diagnosis. They were also asked about story features to be expected from patients with narcissistic and schizoid difficulties and about the difference in stories produced by patient and non-patient subjects.

The tapes were transcribed and inspected for common themes and explicit predictions about what might be found in stories. The following six themes emerged:

1. Borderline features
2. Narcissistic features
3. Schizoid features
4. Engagement with/resistance to the process
5. Pessimism/optimism
6. Violence and aggression

For each theme, at least three statements were framed from the interviewees' predictions. For example, the statements relating to possible borderline features were:

1. Themes of abandonment and being left alone by others are prominent.

2. The main character in this story is passive; the task is achieved by other characters intervening on their behalf.
3. At least one character in this story is sick, ill, poor or in need of rescue.
4. There is at least one character who is a powerless, overwhelmed victim in this story.
5. In this story there is a rescuing, caring character.

It was also desired to test assertions (Lahad, 1992; Lahad & Ayalon, 1993) relating to the BASICPh method of assessment. They maintain that story elements can be found that relate to six domains of coping:

1. Belief
2. Affect
3. Social
4. Imaginative
5. Cognitive
6. Physical

From reading their work, elements relating to each domain were identified and formulated into statements for raters. Again, for each domain at least three statements were identified. For example, statements relating to the social domain were:

- *There are almost no relationships between active participants in this story.*
- *Apart from the main character, there are no other important characters in the story.*
- There are many examples of interactions between characters in this story.

In this example, the two italicised statements would be scored negatively, because they seem to describe a deficit in social coping, whereas the third statement is more positive.

Altogether 50 statements were derived for testing covering the areas described above. The 50 statements about the story were ordered randomly in the response form. This was done by allocating each statement to one of the numbers from 1-50, drawn without replacement by the web-based pseudo-random number generator (Urbaniak & Plous, 2001).

In addition to statements about the verbal content of the story, nine further statements about the visual content of the picture were assembled from the interviewees' tapes and from the researcher's own clinical experience. It was intended that raters should respond to these before they listened to and read the story itself, so in the final rating form these were placed before the 50 statements described above. Examples of these nine statements included:

- The individual images drawn in this picture are very small.
- The drawings seem to have been done with a bold hand, making very positive, definite marks on the paper.
- The picture has obvious images of conflict, violence or death in it.

In addition to the total of 59 statements, there was also a question asking raters to assign a total score on each of the dimensions of Borderline, Narcissism and Schizoid features. They were asked to rate on a 0-10 scale how strongly each of these processes featured in the story in question. These concepts were familiar to raters from the training they had received.

6.3.2.2 Categorical and Qualitative Data

The other main part of the response form was seeking categorical, rather than numerical data. It asked raters to pursue an analysis they had been taught during their training, following a systematic protocol (the Reciprocal Role Grid) to examine the relationships between actants in the story.

The Reciprocal Role Grid involves the rater listing all the actants in a story, then selecting the five most significant and entering them along the top and down the side of a grid. This allows the role each plays towards the other to be listed, with a view to pairing up what A does to B with what B does to A and producing a reciprocal role pair. For example, the following partial grid shows how the story of Cinderella might begin to be coded:

		SUBJECT (Active, doing element of pair)				
		(A) CINDERELLA	(B) STEP- MOTHER	(C) FAIRY GOD- MOTHER	(D) STEP- SISTERS	(E) PRINCE
OBJECT (Passive, done-to)	(A) CINDERELLA		Abuses, ridicules, takes advantage of			
	(B) STEP- MOTHER	Puts up with, submits to, escapes from				

Figure 6.1: Partial illustration of a reciprocal role grid

From practical experience it has become clear that observable, active verbs or verb phrases are the best things to fill into the grid. Not every character has something to do with every other (for example the Prince and Fairy Godmother have no direct relationship). In this example the final reciprocal role pair played out between Cinderella and the Stepmother will look like this:

Abuses, ridicules, takes advantage of



Puts up with, submits to, escapes from

The concept of reciprocal role pairs (RRPs) comes from the practice of Cognitive Analytic Therapy (Ryle, 1990) and is becoming more widely known outside CAT circles as a useful way of describing dyadic relationships. The above example yields a RRP that may be a familiar one to many patients with a personality disorder and their clinicians.

In this pilot, raters were asked to fill in the 5x5 grid twice; once with a free hand to use whatever verbs or phrases they wished to describe the relationships between the actants, as in the example above. The second grid was filled out using the same five actants but a limited repertoire of verbs. These were adopted from the SASB: the Structured Analysis of Social Behaviour (Benjamin, 1996). She identifies 16 positions one individual can take towards another in a dyadic relationship, and proposes that these form an exhaustive list

to which all dyadic relationships in personality disorder can be adequately reduced. The verbs she lists are:

Ignores	Emancipates	Trusts	Controls
Walls off	Sulks towards	Blames	Loves reactively
Discloses to	Separates from	Protects	Attacks
Affirms	Recoils from	Actively loves	Submits to

Each is given a precise operational definition by her, which was provided to the raters in this pilot to help them in identifying appropriate verbs. In the case of the Cinderella - Stepmother dyad above, these might be:

Ignores	(Without giving it a second thought, Subject uncaringly ignores, neglects, abandons Object. Without giving it a thought, Subject carelessly forgets Object, leaves him/her out of important things.)
↕	↕
Submits to	(Subject thinks, does, becomes whatever Object wants. Subject defers to Object and conforms to Object's wishes.)

Finally, each rater was asked for each story to make a judgement about what was being communicated metaphorically. Raters were given three spaces to write in any concept, request or statement that they thought the patient might be communicating through the story. A single final space was given to ask what relevance the rater thought the main character's task in the story might have to the patient's own task in their life or their therapeutic relationship.

6.3.2.3 Data on Raters

In addition to data about the story, pilot raters were asked to give their profession, the number of years experience in that profession, the employing agency and a simple 1-3 rating of subjective confidence in using and rating the 6PSM. At the end of the form, raters were also asked how long it had taken them to complete.

6.3.2.4 Recruitment of Pilot Raters

Pilot raters were recruited from among professionals who had previously attended one of a series of one-day training events in the 6PSM run by the researcher. Eighty-five such participants were approached, of whom 28 agreed to participate. Completed rating forms were eventually received from 19 raters. Table 6.1 below describes those approached, agreeing to participate and actually returning forms.

Table 6.1: Profession of raters in pilot study.

PROFESSION	Approached		Agreed		Returned	
	n	%	n	%	n	%
Art Therapist	14	17	2	7	1	5
Clinical Psychologist	12	14	2	7	2	11
Psychotherapy trainee	12	14	6	21	6	32
Dramatherapist	11	13	7	25	3	16
Nurse	11	13	2	7	1	5
Social Worker	7	8	1	4	1	5
Occupational Therapist	5	6	1	4	0	0
Counsellor	4	5	3	11	3	16
Drama graduate	3	4	2	7	1	5
Lecturer	2	2	1	4	0	0
Charity Worker	1	1	0	0	0	0
Dance Movement Therapist	1	1	1	4	1	5
Music Therapist	1	1	0	0	0	0
Psychiatrist	1	1	0	0	0	0
TOTAL	85	100	28	100	19	100

6.4 RESULTS OF PILOT STUDY

19 raters provided data on the same 2 stories. For statistical analysis of the quantitative data there were therefore 38 'cases', each 'case' being a single rater's rating of one story.

6.4.1 PROFESSION, EXPERIENCE AND CONFIDENCE

Raters' self-assessment of their experience in using the 6PSM was expressed on a three point scale, with the following results:

Table 6.2: Self-assessment of ability to use the 6PSM

Self-rating	Description	No of raters
1	I have had at least half a day's training on the 6PSM but have never used it outside that context.	8
2	I have tried out a 6PSM on other people but I do not feel confident that the results I produce are valid.	4
3	I have tried out a 6PSM on other people and I feel reasonably confident that I can produce some valid results via the method.	7

There appeared to be a difference in the length of professional experience between those expressing more confidence in using the 6PSM and those expressing less confidence. In particular, the median experience in the profession seemed to be very different between the two groups.

Table 6.3: Experience in profession and confidence in 6PSM

Self-rating	Number of raters	Mean years in profession	Median years in profession
1 or 2	12	8.0	4.0
3	7	11.4	13.0

The distribution of ages was different in the two groups: positively skewed for those reporting lower confidence (rating = 1 or 2) and negatively skewed for

those with high confidence (rating = 3). The one-tailed Mann-Whitney U-test was therefore used in preference to a t-test to investigate the difference. There proved in fact to be no significant difference in age between those reporting low confidence and those reporting high confidence. ($U = 41.5$, $Z = -.04$, $p = .97$)

Nor could any association be found between different professional groups, length of experience and confidence in using the 6PSM. These data were not therefore explored further.

6.4.2 PRELIMINARY ANALYSIS OF NUMERICAL DATA

Assessing inter-rater reliability with a large number of raters and a small number of subjects (stories) to be rated is problematic, as most procedures assume 2 raters and as large a number as possible of subjects. The present format had been chosen because of the length of the pilot procedure; it was not felt reasonable to ask volunteer raters to spend the many hours that would have been needed to rate more than two stories. As it was, the mean time reported to rate the two stories was 2 hours 54 minutes (range 1 hour 30 minutes – 6 hours 30 minutes.) The aim at this stage was to reduce the number of statements as far as possible by rejecting those with an obviously poor inter-rater reliability. This would produce a shorter instrument that would be much quicker to complete, with the aim of then getting two more raters to rate a larger number of stories. This would enable final selection of individual items with good inter-rater reliability by more usual means.

Krippendorff (1980) suggests that reliability between more than two raters can be tested by measuring Cronbach's alpha, but this requires scales not individual items. The 59 Likert scale items therefore had to be grouped into scales.

The nine questionnaire items relating to the visual content of the story were analysed as one potential scale first. Streiner & Norman (1995: p65), recommend a value for alpha of between 0.7 and 0.9, and five of the nine items had to be removed before the value of alpha rose above 0.7 (to a value of 0.71). The four visual content items remaining were as follows:

- Every part of this picture seems to be full of rich decoration and detail.
- The picture has obvious images of conflict, violence or death in it.
- This picture has a great many different images drawn in it.
- There is much use of colour made in this picture.

These items were carried forward into the next phase of formal inter-rater reliability testing, described later.

To construct some potential scales from the remaining 50 items relating to the story itself, a factor analysis of the 38 responses (19 raters x 2 stories) was undertaken. Extraction was by the Principal Components method with varimax rotation. This yielded a possible solution of five factors that seemed to group round the following themes. (One item correlating negatively is indicated in italics).

Factor 1: Items describing borderline themes, for example:

- Themes of abandonment and being left alone by others are prominent.

- There is at least one character who is a powerless, overwhelmed victim in this story.

Factor 2: Items describing an impoverished quality of the story itself, for example:

- The story seems to be concrete, straightforward and have little hidden meaning.
- The story seems to be full of metaphor, ambiguity and potential meaning.

Factor 3: Items describing a lack of emotional/relationship description, for example:

- There are almost no relationships between active participants in this story.
- Words describing emotional states are almost never used in this story.

Factor 4: Items relating to negativity or pessimism, for example:

- The story as a whole seems to be pessimistic or negative.
- The whole atmosphere of this story is barren, bleak and lonely.

Factor 5: Items relating to the amount of fantasy material in the story, for example:

- Magic powers, wishes or spells are important in this story.
- The settings, characters and events in this story mostly have magical or fantasy elements to them.

The five factors had the following numbers of items respectively: 15, 12, 8, 9 and 6. Next a scale reliability analysis was carried out on each of the five scales

in turn, eliminating items until Cronbach's alpha was maximised. The number of items removed and the final value of Cronbach's alpha were as follows:

Table 6.4: Final value of alpha for five factors

Factor	No of items removed	Final value of alpha
1	5	0.91
2	0	0.79
3	2	0.77
4	2	0.70
5	1	0.70

In light of the recommendation mentioned above (Streiner & Norman, 1995), these values were felt to be acceptable. 10 items had been removed, leaving 40 still as candidates for the next stage.

The five scales and their items were again inspected, and it was noted that almost all the items in factor 2 required some kind of speculative inference on the part of the rater, for example:

- It seems likely that the main character and task are good metaphors for the writer themselves and their own goals.
- This story seems likely to be a simple retelling of actual events of the teller's own life.

Items on other scales seemed to be more directly about the story itself and to require less speculation on the part of the rater. It was felt preferable to keep all items relating directly to the story itself, and to remove those requiring an inference to be made. All the 12 items relating to factor 2 were removed, leaving 28 items of the original 60. This would leave space in the final instrument for

further items to be derived from inspection of the stories themselves as they were collected.

The other numerical data that were gathered were the raters' 1-10 scores for the presence of borderline, narcissistic and schizoid features in the stories.

These appeared to show some differences in ratings given to the two patients.

Table 6.6: Ratings of features in stories 1 and 2

Categories	Mean rating (standard deviation)	
	Story 1	Story 2
Borderline (n=19)	9.1 (1.7)	6.1 (2.2)
Narcissistic (n=19)	2.3 (2.6)	5.4 (2.3)
^a Schizoid (n=17)	3.7 (2.7)	2.6 (3.5)

^a This rating was not attempted by two of the 19 raters for either story

The Wilcoxon signed ranks test was performed to see how far the ratings given to story 1 differed from those given to story 2 by each rater. The ratings given were significantly different for the borderline ratings ($Z = -3.6, n = 19, p < .001$) and the ratings of narcissism ($Z = 3.4, n = 19, p < .005$). Story 1 was rated as having significantly more borderline features and Story 2 as having significantly more narcissistic features. However there was no significant difference between the stories in the ratings given for schizoid features ($Z = -1.1, n = 17, p = .26$).

This form of rating appeared promising, but depended on an understanding of precisely what was meant by borderline, narcissistic and schizoid. The raters in this pilot had all undergone training with the researcher where these definitions were explored and some common understanding reached. In the main study not all the raters would have had this training. Furthermore, if a rating instrument for general use was to be produced there would be even more variability

introduced by people's different understanding of these terms. It was decided to keep an overall rating scale for the main study but to make it simpler and anchored in concepts that would be easier to define for most clinicians who might use the 6PSM.

6.4.3 PRELIMINARY ANALYSIS OF CATEGORICAL DATA

The raters' first task with the categorical data was to identify the actants in the story and put them in rank order. Story 1, from the patient with Borderline Personality Disorder, will be described to illustrate the process.

Each rater was given space to rank up to ten possible actants in the story. The 19 raters between them identified a total of 17 possible actants. The four actants mentioned most frequently were:

Table 6.7: Highest ranked actants in Story 1

Actant	Mean rank	Mentions
Starfish	1.5	19
Rescuer	1.7	19
Sea	3.7	19
Tide	4.2	17

The mean rank was calculated by averaging the rank position given to each actant by the 19 raters (or fewer, if it was mentioned less often.) There appeared to be considerable agreement therefore between raters, at least in identifying the two most important actants. (A similar picture appeared in the rankings of Story 2, from the patient with OCD.) In fact 17 of the 19 raters had the Starfish and the Rescuer ranked first or second. The remaining two raters had the Rescuer and the Sea in the first two places.

6.4.4 ANALYSIS OF RELATIONSHIPS FROM GRIDS

The next data to be analysed were those from the relationship grids filled in using verbs from the Structured Analysis of Social Behaviour (SASB). One rater was not able to fill in the grids for either story, so figures henceforward relate to 18 raters, not 19.

The SASB roles assigned to the actants ranked first and second were inspected. These would form the most significant reciprocal role pair and the relationship between them was presumably an important one. Could raters agree on the description of this relationship? The reciprocal relationships identified by the 18 raters were as follows:

Table 6.8: Relationships identified by raters – Story 1

<u>Reciprocal Relationship</u>	<u>No of raters</u>
Protects – Trusts	8
Emancipates – Trusts	3
Actively loves - Trusts	2
Protects – Submits to	2
Loves reactively – Submits to	1
Protects – Loves reactively	1
Controls – Walls off	1

Given the 16 verbs available in the SASB, there are $15 + 14 + 13 \dots + 1 = 120$ different RRP combinations possible. Only 7 of these 120 possible combinations were mentioned at all, and of the 18 raters, eight agreed on a single one of the 120 combinations. The Protect-Trust dyad is certainly one that might be predicted in someone with a Borderline Personality Disorder, and in this particular patient's case it was a very relevant clinical feature. This was the kind of relationship that the patient was constantly seeking but never finding.

However looked at in another way, fewer than half (8 out of 18) of the raters could agree on the same description of the most significant relationship. Raters' performance on Story 2 was no better. Additionally, in their feedback raters had described that this process of constructing the role grid and identifying the reciprocal roles was by far the most confusing and time-consuming element. It was therefore decided that, however clinically interesting this process might be, it was insufficiently rigorous and replicable for a research project. This method of rating was therefore omitted from the next phase of the study.

7.0: INTRODUCTION TO THE MAIN STUDY

This chapter describes the method used for recruiting participants, as well as the protocols for eliciting and managing data. In practice, some of the work described in this chapter overlapped with that described in the previous Chapter on the pilot study. It is presented here separately for the sake of clarity.

7.1 OVERALL STUDY DESIGN

The study was designed to see whether the 6-part story method could be used to identify any differences in the stories produced by participants with and without a diagnosis of Borderline Personality Disorder (BPD). The working hypotheses were these:

1. Raters who were blind to the authorship of 6-part stories would be able to agree between themselves on statements rating these stories.
2. Some of the ratings given would cluster into psychologically meaningful groups of statements such that scores could be calculated.
3. Scores given to stories from individual authors would show a degree of stability over time (an interval of one month between stories recorded at two time points.)
4. Scores given to stories would correlate with some of the concurrent clinical data collected from patient participants.
5. Computer-based text analysis would reveal differences in language use between stories from those with a BPD diagnosis and others.
6. Some of these differences would show acceptable levels of test-retest reliability and concurrent validity.

7.2 ESTABLISHING A PROTOCOL FOR ADMINISTERING THE 6-PART STORY METHOD

The author had been teaching the 6PSM to groups of clinicians for the previous five years. Part of the teaching package included a script for the administration and discussion of the 6-Part Story, which clinicians were encouraged to customise and make their own once they were familiar with it. Questions to the story teller were given as examples only, not as prescriptive scripted prompts. For the purposes of this research a more standardised method of eliciting the story was required, in order that all tellers received as nearly as possible the same prompts.

A new script was written, with the intention that it would be read word for word by the researcher to each clinician subject as part of their training in the 6PSM. The same script would then be used by each clinician subject to elicit the 6-Part Stories from every patient subject. Questions about each part of the story were drafted, being left sufficiently open-ended and general that they could be asked of any teller about any story – for example:

Staying in the first picture, tell me about the place a bit more. Tell me a bit more about where we find the main character; what kind of place is it?

The script was piloted on the first groups of clinicians to be trained in the 6PSM and their reactions to it elicited. Only small amendments were needed, the principal one being that the number of supplementary questions being asked was reduced as it became clear that some of these were repetitive and redundant. The layout of the script was also amended to make it more readable.

Clinicians reported that it felt odd reading a script rather than putting the instructions into their own words. However, when telling a story and having instructions and questions read from a script they reported that from the storyteller's point of view this artificiality was not intrusive.

The end result was a script and set of instructions for administering the 6PSM which are attached as an Appendix 3.

7.3 RECRUITMENT OF PARTICIPANTS

A target of 30 clinicians was set, to represent the geographical spread and multi-disciplinary makeup of CMHTs in a single NHS Community Trust. Stories were recorded from Community Mental Health Team (CMHT) clinicians as part of a training session to equip them to record stories from patients on their own caseloads.

Each clinician then identified, with the researcher's help, two patients from their caseload. One was someone meeting three outline criteria for a possible diagnosis of BPD, and the second was a more mainstream CMHT client meeting none of these criteria. The aim was for clinicians to record two stories from each patient, with a one-month gap. Between the two story sessions, the researcher met each patient for a single session to undertake a structured clinical interview and administer two self-report questionnaires. The aim was therefore ultimately to have 30 stories from 30 individual clinicians, 60 stories

from 30 patients with a BPD diagnosis, and 60 stories from 30 patients without such a diagnosis. The actual numbers achieved will be described later.

The stories were transcribed and copies of the transcripts sent to raters who had no other information given to them about the stories or their authorship. Raters filled out a rating form using the statements whose development has already been described and returned the transcripts and ratings to the researcher for analysis.

In addition to the quantitative analysis for which these data were collected, every participant was also asked after their storymaking session for their immediate reactions to the process. These were recorded and transcribed, but were not made available to the blind raters described above. These reactions were collated for qualitative thematic analysis by the researcher.

7.4 ETHICAL APPROVAL

This was sought from the Local Research Ethics Committee in application dated 5 November 2001. Final ethical approval was given by the committee on 2 January 2002. The patient information sheet and consent form approved by the LREC are reproduced at Appendix 2.

7.5 TRAINING OF CLINICIANS

Clinicians working in an adult community mental health service of an NHS Community Trust were approached. Social Services staff working in the adult

mental health resources within the Trust boundaries were also approached. All teams (9 NHS, 9 Social Services) were sent a letter outlining the study and interested individuals were asked to contact the researcher. A target of 30 clinicians was aimed for and this number was reached by July 2002.

Clinicians were trained in groups of 2 or 3. Each group lasted 2-3 hours, and started with a recap of the information in the letter and the chance for questions. All clinicians were happy at this stage to agree to participate and the next stage of the session was to teach the 6PSM. The researcher read out the first part of the script to the clinicians, who drew the pictures as instructed and began to formulate a story. The researcher then passed the script to one of the clinicians, who continued to read it out. This started from the point where the storyteller is asked to tell their story through once without interruption, and continued through the question and answer phase to the end. This part of the session was tape recorded in every case. The clinicians then swapped roles until everyone had acted both as storyteller and as listener. Feedback was given by the researcher on the level of consistency reached and on the decisions made as to whether or not to use supplementary questions. Most training sessions ended with a short free conversation between all participants about their experience of undertaking the 6PSM and their hypotheses about how, if at all, it was working. These conversations were recorded on a separate tape.

7.6 SELECTION OF PATIENTS

Where there was time, patients were selected within the training session.

Where this was not possible a short further session (face to face or telephone)

was arranged. Clinicians were reminded of the criteria for choosing two patients. The first patient was to be one who met the simple criteria for referral to the researcher's personality disorder team, namely:

- Long history of crises where some form of self-harm (e.g. cutting, overdosing) was a feature. Repeatedly threatening self harm or suicide would also count. Clinicians were also asked to consider less obvious forms of self-harming behaviour, such as eating disorder, alcohol or drug use (when not a purely addictive behaviour), sexually risky behaviour etc.
- If there had been psychiatric admissions, these would tend to be chaotic, unplanned and frequently break down unproductively
- Multiple psychiatric diagnoses and treatments, or disagreement or uncertainty about diagnosis and treatment within the team. For example the patient might have been described formally in notes or informally in team discussion as 'not really mentally ill', 'manipulative' or 'behavioural'.

These criteria, although apparently crude, had been found in practice to be very easy for front-line practitioners to apply. They also appeared to distinguish well between patients with a personality disorder (particularly Borderline Personality Disorder) and those without. Dunn & Parry (1997) describe how the criteria were arrived at and used in practice in the personality disorder service being described.

The second patient was to be one who met none of the above criteria, in other words:

- No repeated history of self-harm or threats of self-harm (one or two occurrences would not exclude, if these were seen as atypical for that patient.)
- If there had admissions these would tend to be productive, with planned discharges
- Well understood diagnosis with agreed treatment plan

Clinicians were told that these criteria did not preclude the fact that the second patient might have a serious mental health problem. For example, somebody with a serious manic depressive or psychotic disorder could be included provided they met none of the criteria for the first patient.

Each clinician was asked to make a list of their patients who fell into one or the other group. In allocating patients to a group, clinicians were also asked to make a clinical decision about whether they would, in principle, be happy to approach the patient to take part in the study. They were asked to apply the following criteria when excluding a patient from consideration:

- Patients younger than 18 or older than 65
- Patients whom the clinician considered to experience profound communication difficulties.
- Patients who were currently psychiatric inpatients
- Patients with psychotic symptoms that were not currently well controlled
- Patients currently receiving formal psychotherapeutic treatment
- Patients whom the clinician would not be willing to be shown a short report produced by the researcher after the assessments

- Any other patient whom the clinician felt it unwise, for clinical reasons, to approach.

A belief on the part of the clinician that the patient might refuse, would find the process difficult or would be unlikely to produce a useful story was NOT sufficient to exclude a patient. Clinicians were encouraged to include as many patients as possible into their two pools and not simply cherry-pick those they thought might be co-operative or productive subjects. Typically, for each clinician there were 2 or 3 patients in the pool of those with a possible personality disorder (a few clinicians had as many as 10) and 10-30 in the mainstream CMHT pool. The 30 possible PD patients were therefore to be randomly selected from a total pool of about 75 patients, and the 30 mainstream patients from a pool of about 600.

A random number table was then used to select the two patients, one from each pool, to approach. A fall-back patient from each group was also selected, in case the first patient did not agree to participate. Initially an attempt was made to produce pairs of patients matched for gender and age (± 10 years.) This was impossible within the patients of a single clinician because of the small numbers in each clinician's possible PD patient pool. Trying to match patients across clinicians also proved very difficult because of the numbers of patients who declined to take part, or who initially agreed but later changed their mind. It was reluctantly decided to take any patient who met the outline criteria, and to look at the distribution of age and gender produced once the groups had been selected.

7.7 DATA GATHERING

7.7.1 DATA FROM FIRST STORY SESSION

Clinicians sent the following items back to the researcher after recording their first 6-Part Story with each patient: completed patient consent form, original of picture drawn by patient, original tape recording of 6-Part Story. At this stage the researcher did not know whether the individual was one with a potential personality disorder or a more mainstream patient. The tape and picture were not inspected by the researcher at this stage.

7.7.2 DATA FROM INTERVIEW WITH RESEARCHER

The next step was for the researcher to meet the patient to conduct the SCID-II diagnostic interview and to give the patient the CORE and IIP-32 questionnaires. These meetings were arranged through the clinician, and took place in whatever setting the clinician usually saw the patient. In almost all cases this was the Community Mental Health Team base, though in a small number the patient's home was used.

The reasons for choosing the SCID-II have already been discussed in Section 5.5.1. Unfortunately it was not possible to locate any formal training in the SCID-II in the UK, despite considerable effort. The researcher had however undertaken training in the similarly-structured International Personality Disorder Examination developed by Loranger and published by the World Health Organisation (1995). The SCID-II was preferred to the IPDE because of the

existence of a screening questionnaire for the SCID-II which would speed up administration of the interview. All patients were given a copy of the questionnaire by their clinician at the close of their first story-making session, and these were then brought by the patient to the interview session with the researcher.

The interview began with the researcher giving the patient the CORE questionnaire to complete while the SCID-II questionnaire was scored. This simply involves identifying any of the 104 questions the patient has answered yes to (or left blank); only these items are discussed in the SCID-II interview proper. This is because it is assumed that the questionnaire items produce far more Type 1 errors than Type 2. In other words patients saying 'no' to a problem are not likely in fact to have it, and these questions can be ignored. Patients who say 'yes' on the other hand will often be over-reporting problems, and these need to be probed by the researcher.

During the interview, the researcher made decisions about whether or not any criterion had been met as the patient answered the prompt questions given in the SCID-II handbook. In cases of doubt, questioning on an item continued until the researcher was able to code the patient's response. No coding was left to be done retrospectively in the absence of the patient after the end of the interview.

Finally, once the SCID-II interview was over, the patient was finally given the IIP-32 questionnaire to complete and asked some brief questions about age, living arrangements and level of education reached.

The interview session with the researcher took between 70 and 135 minutes. At the end of the session they were reminded that their clinician would be recording a second 6-Part Story with them soon, and that following receipt of this a copy of their picture, tape transcripts and questionnaire results would be returned to them by the researcher.

7.7.3 DATA FROM SECOND STORY SESSION

Approximately four weeks after their first storymaking session, clinicians were asked to record a second 6-Part Story with each patient. As with the first story, the picture and tape were sent to the researcher. Once both stories were obtained from the clinician's two patients (possible PD or mainstream CMHT), only at this stage did the researcher ask the clinician which patient was which. This was to prevent, as far as possible, the researcher from knowing the group membership of the patient at the time of the SCID-II interview.

7.8 RECRUITMENT AND TRAINING OF ADDITIONAL RATERS

It was planned that the clinician participants in the study would also rate stories (other than ones they had gathered). However other raters were also approached, to increase the number of ratings obtained. The other raters were drawn from the courses in the 6-part story method that had been run by the researcher over the previous three years in various parts of the country. Both the raters recruited in this way, and the clinician participants from the CMHTs received similar training. The external raters had all received a somewhat

longer training (one working day) than the CMHT clinicians (approximately two hours) but the same essential element of learning the process by creating and telling their own story was common to both groups. Neither group was given any specific training in the rating methods to be adopted. Any instructions in the rating method were given purely in writing on the rating forms themselves.

8.0 PRIMARY DATA: COLLECTION AND PROCESSING

The primary data were those gathered from and pertaining to the participants in the study, both clinicians and practitioners. All participants produced 6-part story pictures and tapes, while patient participants also provided clinical interview and self-report data. This chapter describes how these data were gathered and handled. Chapter 9.0 will describe the secondary data: the ratings and other data pertaining to the stories and produced by the panel of raters to whom story transcripts were sent.

8.1 INITIAL PROCESSING OF DATA

8.1.1 PICTURES

Pictures were labelled on the back with a unique 3-figure code, with the same code being used to identify the tape and the transcript. Patients and clinicians had sometimes put a name, date or initials somewhere on the picture; where this happened such identifiers were obscured by correction fluid and a white label stuck over them to make them unreadable.

8.1.2 TAPES

Tapes were transcribed from the point where the patient begins to tell their story to the end of their response to the last scripted question. The clinician's questions and prompts were not transcribed directly; rather a template was used from the given script and the patient's replies typed into the template.

As transcription started it became clear that transcribing extemporised, spoken stories into written text was problematic. Spoken stories frequently lack obvious sentence structures, and all punctuation must be inferred. Some tellers used dialect forms that might betray their level of education. Others frequently departed from telling the story to ask for reassurance, comment on the process or for other reasons.

It was decided to make the transcribed texts as nearly like one another as possible, in order that clues from repetitions, dialect forms and self-corrected mistakes should be minimised. To this end a style sheet for transcription was developed which was followed throughout. To ensure adherence to the style, all transcription was carried out by the researcher. Style sheet examples include:

Table 8.1: Words to transcribe more formally

Word used by patient	Word transcribed
'Cos	Because
Nowt	Nothing
'Til	Until
Dunno	Don't know
There wasn't no...	There wasn't any...

Table 8.2: Word combinations to remove

Word/s to remove	Under what circumstances
Erm, er, ah	And all similar non-word utterances
Sort of, kind of	Unless used purposefully to mean a kind or variety
You know	When used as a filler
I think, well, so,	At start of sentence as filler, or as a type of pause marker
Repetitions	Unless deliberately used for emphasis
Questions	to interviewer "What was the third picture?", or repeated to self "What is the atmosphere like?"
Observations	about the process – eg "This is difficult, isn't it?", "It's a daft story", "I'm not very creative", "I don't know"
Requests	For guidance or reassurance: "Is this OK", "Am I doing it right?"
I guess, I suppose	Unless in character's reported speech
just, really, like, quite	When used as empty adjective. (Leave in if it is obvious that just = simply, really = very, like = similar to)
That kind of thing, and so on, etcetera	When appearing at the end of a list

Simple mistakes in grammar were corrected, for example “He were going to try and escape” was transcribed as “He was going to try and escape”. Sentences were formed when it seemed that a complete meaning unit had been reached and a new one started, with punctuation inserted to make reading the text as easy as possible for the reader. Where more complicated corrections would have been needed to make the sentence grammatical, or where the train of thought apparently became confused, no attempt at repair was made. The text was transcribed in a way so as to make it as close as possible a representation of the words as spoken. For example:

Clinician: In the third picture we have the things that make it harder for the main character. Can you tell me a bit more about the things that make it difficult?

Patient: Because it is the brick wall, but he can't get it, there is a brick wall, the key is through the brick wall.

Occurrence of onomatopoeic words (splash, boom, boing) was rare, and there was fortunately no use by anyone of sounds which could not be transcribed (eg raspberry blowing, handclap or similar non-verbal sounds.)

The very last portion of the tape, where the patient is asked to comment on the storymaking process and their reactions to it, was transcribed but into a separate document, so that all these reactions could be analysed together. Computer and human rating of the stories was done on the rest of the tape, up to the point at which these questions started.

8.1.3 SCID-II INTERVIEW

For each of the 12 possible DSM-IV personality disorder diagnoses, the total number of criteria met was counted and entered into a spreadsheet. When the minimum number of criteria for the presence of the disorder were met, the spreadsheet was set up to calculate and record this as a dichotomous variable (yes/no.) The total number of DSM-IV criteria met and the total number of personality disorders identified for each patient were also calculated by the spreadsheet and recorded. The MS Excel spreadsheet was then read directly by SPSS for statistical calculation. All the data manipulation from the entering of the raw number of criteria met was therefore done automatically, to reduce as far as possible the chance of transcription or calculation errors.

8.1.4 CORE AND IIP-32 QUESTIONNAIRES

These questionnaires were scored according to their respective manuals and the raw scores for each subscale added up by hand. The subscale scores were then entered into a spreadsheet where the population mean and standard deviation from the normative data were also present. For both CORE and IIP-32, a 'normal' and a 'patient' mean and s.d. are given in the normative data. For this study the patient subjects were compared against the 'patient' norms by using their raw scores to calculate a z-score for subscales and main scale. This allowed comparison between male and female patients, for whom the norms were slightly different. In the case of one patient who was a male to female transsexual, it was decided to use the female norms to calculate their z-score.

The z-score calculations were carried out within the spreadsheet and pasted into the same file as data from the SCID-II interview, rather than being copied and typed in.

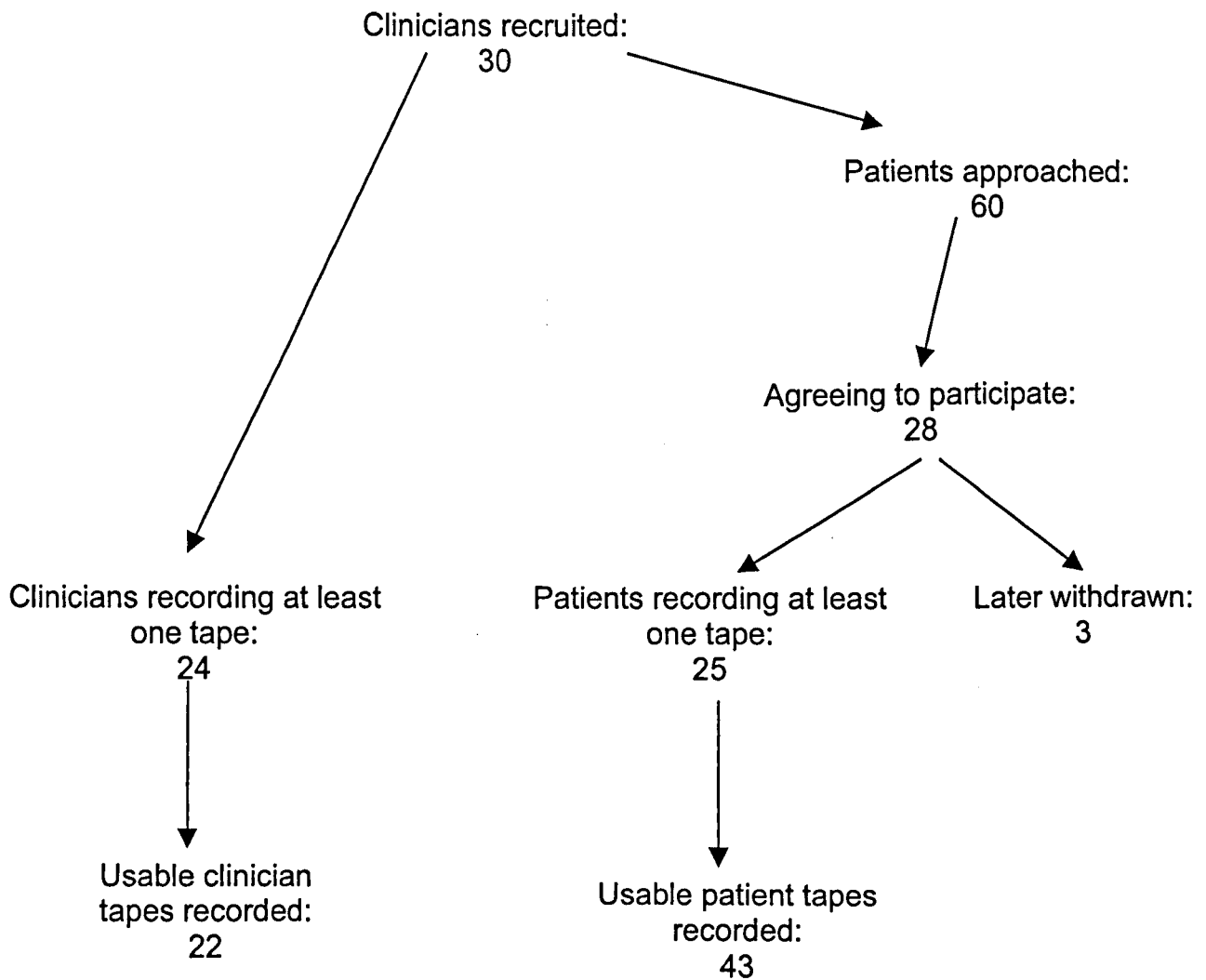
8.1.5 OTHER DATA

Gender, age, living situation and other data were typed directly into the master spreadsheet file, along with codes to identify the patient, their clinician, the code numbers from their two stories and any other relevant data for processing.

8.2 CHARACTERISTICS OF THE SAMPLE

Once the data gathering process described in Chapter 4 was complete, 65 story transcripts produced by 49 individuals had been obtained. Each individual produced either one story or two. The 49 individuals comprised 24 clinicians from the mental health service and 25 patients from their community caseload. There were in addition three stories from three further patients who subsequently asked to withdraw from the study. Their stories were not rated or included in the quantitative analysis in any way. A summary of the recruitment process is given graphically below:

Figure 8.1: Recruitment path of participants



65 stories was less than the planned-for 150 transcripts from 90 individuals. As it became clear that fewer stories than planned were being collected, CMHTs and participating clinicians were contacted again to ask if they had any further patients who would be willing to contribute to the research. In order to make the burden on clinicians as light as possible and encourage them to recruit further clinicians, the plan for data gathering was changed with the researcher undertaking the taped story sessions as well as the intervening diagnostic interview. This change was made by agreement with the Chair of the Local

Research Ethics Committee. As a result two further patients, recording two stories each, were recruited.

As described earlier, the CMHT patients were in two groups; those whom their key workers had identified as having a possible personality disorder and those identified as more mainstream CMHT patients. The numbers of stories from these groups were as follows:

Table 8.3: Number of participants and stories from each group

Group	Individuals	Stories
Clinicians		
Recording 1 story	24	24
Patients: possible PD		
Recording 1 story	4	4
Recording 2 stories	7	14
Patients: Mainstream		
Recording 1 story	3	3
Recording 2 stories	7	14
Patients: Group not recorded		
Recording 2 stories	4	8
Total	49	67

The majority of the 49 authors were female, among both clinicians and CMHT patients, as shown in Table 8.4 below. The most obvious difference between the patient groups is that the possible PD group has a majority of women in it while the mainstream CMHT group has a majority of men. This is a significant difference ($\chi^2 = 5.74, df = 1, p < .05$). The ages of the client groups were comparable, as was the proportion living with a partner and the proportion having post-16 education. Apart from gender, these data were not available for the clinicians. However it can be assumed that all of them had post-16 education as all were professionally qualified.

Table 8.4: Demographic details of participants, by group

Group	Gender		Total	Mean (sd) age ^a	Proportion living with partner ^a	Proportion with post-16 education ^a
	Male	Female				
Clinicians	4	20	24	n/a	n/a	24/24
Patients: Possible PD	2	9	11	34.9 (8.3)	5/10	4/10
Patients: Mainstream	7	3	10	36.1 (11.0)	3/9	4/9
Patients: Group not recorded	0	4	4	35.8 (12.3)	0/4	2/4
Total	13	36	49	35.5 (9.6)	8/23	34/47

^a Data in these columns are incomplete because two patients who recorded stories were not interviewed. The totals in these columns do not therefore match those in the fourth column.

Once the SCID-II data were available, it was possible to look at the actual personality disorder status of the patients, rather than their status as predicted by their key workers. The criteria for membership of the possible PD group had been drawn up to maximise the chances of identifying, in particular, patients with a diagnosis of Borderline Personality Disorder (BPD). In this, the crude criteria given to key workers seemed to be very successful, as seen in Table 8.5 below:

Table 8.5: BPD diagnosis as predicted by key worker vs. actual diagnosis

Predicted BPD diagnosis	Actual BPD diagnosis		Total
	No	Yes	
Possible PD	1	9	10
Mainstream	9	0	9
Not recorded	2	2	4
Total	12	11	23

Note: Two of the 25 patients declined to undertake the clinical interview that would have produced the BPD diagnosis.

Of the 19 patients where the group membership was recorded, only one was apparently misallocated by their key worker. This was a patient who was allocated to the possible PD group by her key worker, but who proved to have no PD detectable by the SCID-II. The key worker felt that the patient in question was self-harming through her use of food rather than more obvious means such

as cutting and overdosing. This was a very experienced key worker with an MSc in psychotherapy and UKCP registration, so it is possible that the SCID-II process was at fault here in failing to detect an actual PD, rather than the fault being with the key worker. However, for the purposes of the study the SCID-II was treated as the gold standard and this patient was subsequently treated in the data analysis as not having BPD.

Looked at in more detail, the SCID-II results for the 23 patients who undertook it are outlined below. It is noteworthy that there is no clear dichotomy here between two groups, despite the sample being generated in this way. Only six of the 23 patients had no Axis II diagnosis at all; five of the nine patients identified as mainstream did, in fact, have at least one personality disorder. This is in line with the findings of Keown, Holloway & Kuipers (2002) who found that 52% of the patients of a south London CMHT had a diagnosis of at least one personality disorder.

Among those with no PD identified there was also wide variation; one of these patients did in fact meet 27 individual SCID-II criteria without crossing the threshold for the diagnosis of any one disorder. Compare this with a patient diagnosed with three Axis-II disorders who only met 24 criteria. These details are given more fully in Table 8.6.

Table 8.6: Axis II diagnoses of individual patients

Patient	Histrionic PD	Dependent PD	Narcissistic PD	Schizotypal PD	Obsessive/ Compulsive PD	Schizoid PD	Avoidant PD	Passive/ Aggressive PD	Antisocial PD	Paranoid PD	Depressive PD	Borderline PD	TOTAL no of PDs
1													0
2													0
3													0
4													0
5													0
6													0
7												X	1
8								X					1
9							X						1
10							X					X	2
11											X	X	2
12								X		X			2
13							X				X	X	3
14								X			X	X	3
15						X				X	X		3
16				X		X			X	X			4
17								X	X	X	X		4
18				X		X					X	X	4
19					X	X			X	X		X	5
20							X		X	X	X	X	5
21			X					X	X		X	X	5
22					X				X	X	X	X	5
23		X			X		X	X	X	X	X	X	8
Cases of PD:	0	1	1	2	3	4	5	6	7	8	10	11	

As expected, Borderline PD was the commonest diagnosis, occurring in 11 of the 23 patients in the group who completed a SCID-II interview. However Depressive PD was nearly as common, and overlapped almost completely with the BPD group. Only three patients with BPD did not also have a diagnosis of co-morbid Depressive PD. In fact there were high levels of co-morbidity for many PD diagnoses. Hereafter the group of patients with a diagnosis of BPD will be frequently compared to the non-BPD patients, but it must be remembered that what is being discussed is actually a group of 11 people, among whom eight also have Depressive PD, five Antisocial PD, four Paranoid

PD, four Avoidant PD, three Passive/Aggressive PD and three Obsessive/Compulsive PD.

Conversely, some PDs were diagnosed among few or no patients. It was perhaps particularly surprising that only one patient had a diagnosis of Dependent PD when this might have been expected to be more commonly associated with Borderline PD than, for example Antisocial PD. It is also interesting that another Cluster B diagnosis, Histrionic PD, was not found at all when it might have been expected to co-occur at least sometimes with BPD. These low figures are in line with a large national sample investigated by the Office for National Statistics in 2000 (Singleton, Bumpstead, O'Brien, Lee, & Meltzer, 2003). This found that the four most infrequently occurring PDs in a general household sample were Dependent, Histrionic, Narcissistic and Schizotypal PDs, all of which had a point prevalence of one per thousand or less.

9.0 SECONDARY DATA:
REFINEMENT AND COLLECTION

The previous chapter described the acquisition of the primary data: interview and self-report data from patient participants and raw story material from all participants. This chapter describes how the raw primary material of story transcripts was transformed into secondary data by a panel of raters who read the transcripts and provided ratings of the stories.

At this point the individual items in the rater response form had still only been tested out by 19 raters on two separate stories. Before the final group of raters were sent the final version of the rater response form for the full study, it was necessary to do some further work on the inter-rater reliability of the individual items.

9.1 OBJECTIVES AND DESIGN

The objective of this stage of the study was to expand on the 28 rating statements identified by the pilot study described in Chapter 6. As described there, with only two stories being rated it was not possible to be certain that any of the statements had adequate inter-rater reliability, and since the two stories came from different patients, test-retest reliability was not tested in any way. The design of this stage was first to test the inter-rater reliability of individual statements by having two experienced raters rate a sub-sample of 20 stories. Following this a larger panel of more naïve raters gave ratings to all the stories in the study, in order to further test the inter-rater reliability of individual statements. Statements were then assembled into meaningful scales that would permit test-retest reliability testing to be carried out. Finally any scales that

passed all these tests were compared against the concurrent clinical data to assess their criterion validity.

9.2 METHOD

Once the main study started, primary data began to be collected in the form of stories from patients and responses from their SCID-II structured interviews and CORE and IIP-32 self-report questionnaires. Once data had been collected from 20 patients these were inspected to see what patterns emerged. In particular, responses from the CORE and IIP were compared with SCID responses to see which scales from the self-report questionnaires might be associated with a borderline diagnosis from the SCID. This was measured by taking the number of SCID criteria for a diagnosis of Borderline Personality Disorder (BPD) that were met for each patient. This number was correlated with all the scales and subscales of the CORE and IIP.

The point of doing this was to identify two groups of stories: one from patients reporting low levels of problems on the concurrent measures, and one group from patients at the other end of the continuum. Stories from the two contrasting groups would be compared to see whether this comparison yielded further statements that could be tested for inter-rater reliability by a larger rating panel.

It was found that the CORE Risk subscale and the mean for the whole IIP were most strongly correlated with scores on the SCID Borderline criteria. These two scales were plotted against each other as a scatterplot, shown in Figure 9.1.

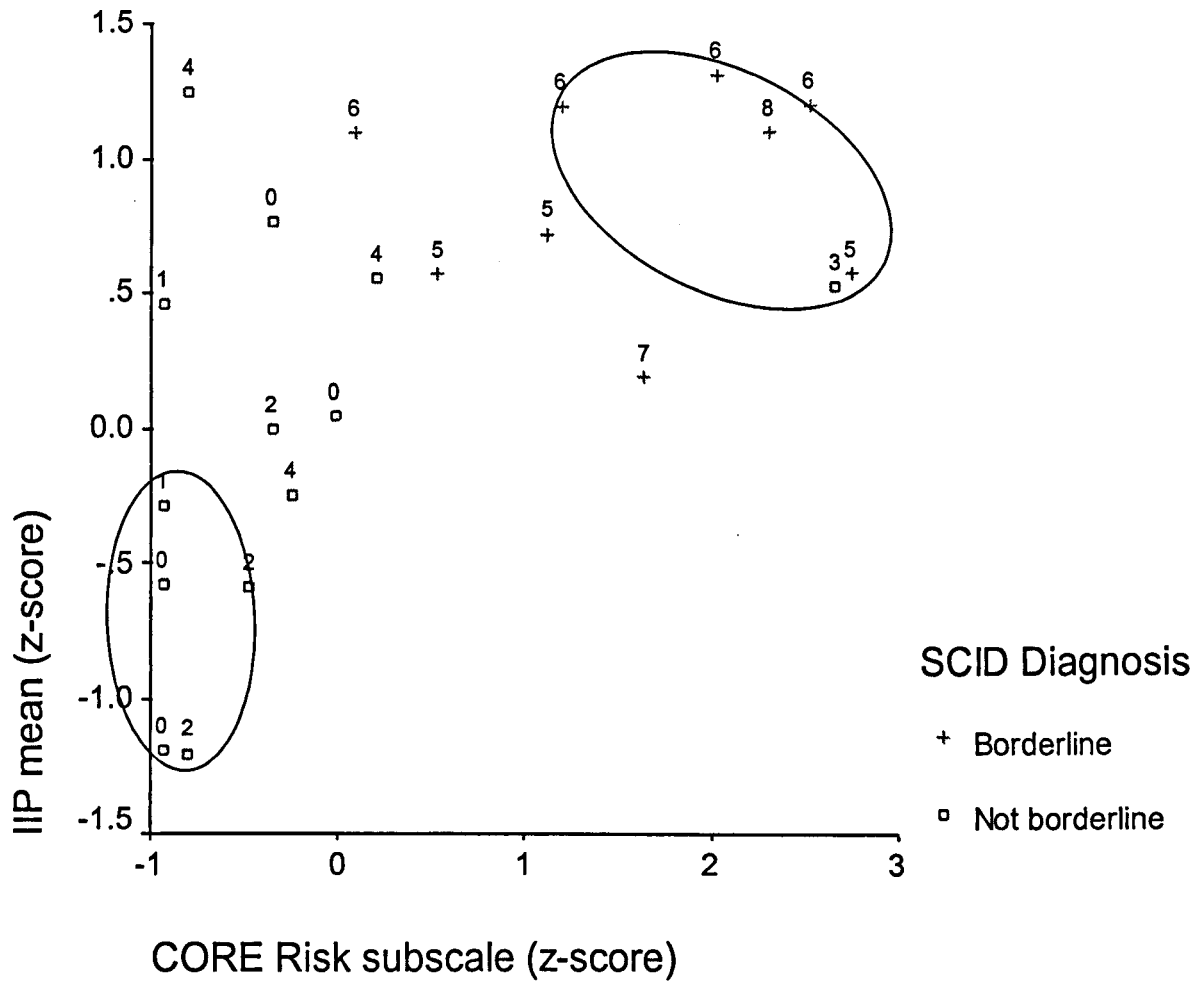


Figure 9.1: Scatterplot of first 21 patients: scores on CORE risk subscale versus IIP mean. The small number above each data point is the number of BPD criteria met by that patient.

The x-axis represents the patient's score on the CORE risk subscale, six questions relating to the risk of harm to self or others. As might be expected, patients with BPD scored much more highly; indeed all scored above the patient mean ($z = 0$) described by the authors (CORE System Group, 1999) in their normative data.

The y-axis represents the patient's mean score on the IIP-32. The normative data used for calculation of these z-scores were those described by Barkham,

Hardy & Startup (1996). Patients with a BPD diagnosis all scored above the normative mean, but so did about half the patients without a BPD diagnosis.

Next, the five most extreme patients from each group (BPD and non-BPD) were chosen, as indicated by the ovals marked on Figure 9.1. The anomalous result of the non-BPD patient occurring at the extreme end of the BPD group was ignored for the time being. Extreme subjects were chosen because it was felt that if there was any difference in the stories produced by the two groups this would be most clearly seen at the extremes.

These ten patients had produced 16 stories between them at the time of this analysis (6 had produced 2 stories each, 4 just a single story.) The 16 stories were read and for each story a list of themes or characteristics was made. As themes emerged in later stories, earlier stories were re-inspected to see whether the theme could apply there also. This process of constant comparison was adapted from that described for analysing qualitative data described by Strauss & Corbin (1998). Once all 16 stories had been inspected in this way those themes and characteristics were selected that seemed to:

- Describe differences between the two groups
- Describe similarities within each group

This process yielded a further 19 items to add to the 28 already obtained in the first step. Items derived in this second process gathered naturally into groups relating to different stages in the six-part story – some to the main character, some to the task and so on. This structure was adopted for the whole of the 47

items now in the draft final instrument, with the order within groups being determined randomly. The final 47 items were as follows:

DRAWN IMAGES

- 1 Every part of this picture seems to be full of rich decoration and detail.
- 2 The picture has obvious images of conflict, violence or death in it.
- 3 There is much use of colour made in this picture.

MAIN CHARACTER

- 4 In this story there is a rescuing, caring character.
- 5 The main character is specifically described as average, ordinary, normal.
- 6 The main character has likeable, admirable qualities
- 7 Some characters in this story are superior, grandiose, of high rank, celebrated or admired.
- 8 There is at least one important character who is larger than life – a god, superhero, film star, famous historical character or similar.
- 9 There is at least one character who is a powerless, overwhelmed victim in this story.
- 10 There is an obvious bullying, dominating, violent character in this story.
- 11 At least one character in this story is sick, ill, poor or in need of rescue.
- 12 Characters in this story seem to have belief and confidence in themselves.

TASK

- 13 The main task is of self-development, satisfying curiosity, enlightenment, achieving potential.
- 14 The main task is about a journey or escape from a hostile or unpleasant situation.
- 15 Failure to complete the task would mean starvation or death for the main character.
- 16 The main task is to capture, kill or eat some other creature or being.

OBSTACLES AND PROBLEM-SOLVING

- 17 Violent, physical means are used to overcome obstacles.
- 18 Opposition makes the main character angry, frustrated...
- 19 Problems in this story tend to be about physical size and strength (or smallness/weakness.)
- 20 Obstacles are overcome with faith, persuasion, skill or problem-solving.
- 21 The main character heroically overcomes stronger opponents.

RELATIONSHIPS AND HELP

- 22 The characters in this story show no awareness of one another's needs and give one another no consideration.
- 23 Others are mostly seen as helpful, positive, friendly.
- 24 Others are mainly a threat towards the main character in the story.
- 25 There are many examples of interactions between characters in this story.
- 26 Themes of abandonment and being left alone by others are prominent.
- 27 Other characters essentially complete the task for the main character.
- 28 Others are helpful because their presence means the main character is not alone.

FINAL OUTCOME

- 29 The outcome is a 'win-win' situation for main character and most others.
- 30 The outcome is positive for the main character.
- 31 The outcome is positive for the main character, but only at the expense of most others.

32 The outcome is negative for the main character.

MORAL OR TEACHING

33 The story teaches that confronting fears and problems is necessary to happiness.

34 The moral is that we all have untapped potential in us.

35 The moral of the story is about persistence, brains or bravery achieving success.

36 The story teaches that we always need other people.

GENERAL THEMES

37 The story as a whole seems to be pessimistic or negative.

38 Positive images of life, growth, health or production predominate.

39 Magic powers, wishes or spells are important in this story.

40 There are many mythological creatures or themes in the story.

41 The whole atmosphere of this story is barren, bleak and lonely.

42 Morbid themes of death, aggression, pain or decay predominate.

43 The settings, characters and events in this story mostly have magical or fantasy elements to them.

44 Themes of good and evil, right and wrong are important in this story.

45 The content of the story is minimal, stark and brief.

46 The story as a whole seems to be an optimistic or positive one.

47 Things that harm others are important to the story – guns, swords, teeth, claws.

9.2.1 ITEM SUITABILITY AND INTER-RATER RELIABILITY

This was tested by having two raters rate a number of stories. Two raters who were familiar with the 6PSM were used. The first rater was one of the clinicians who had gathered 6PSM data from their patients, and who in addition was already very familiar with the 6PSM. The second rater was the researcher, who was not blind to the author of the story and the data gathered from the clinical interview and self-report questionnaires. 20 stories were selected for rating; the 16 patient stories that had been used in the process just described, along with four stories from randomly chosen clinicians.

(Amended section)

Each item was rated on a 5-point Likert scale (strongly disagree – strongly agree). Agreement between raters was measured by calculating the weighted Kappa statistic. This was used because the data were ordinal Likert scale

responses and some credit needs to be given for a near agreement. For example, if rater A rates an item 'Strongly Agree' and rater B gives it 'Agree Somewhat', the regular Kappa statistic would treat this as a disagreement. In other words rater B's response would be regarded just as if it were 'Strongly Disagree'. Weighted Kappa on the other hand gives partial credit for near misses, while remaining sensitive to one rater consistently rating higher than another. This is a problem for another method of analysis considered, the Gamma statistic described by Siegel and Castellan (1988). This was the statistic initially selected, but weighted Kappa was ultimately preferred following advice from a statistician.

9.2.2 FIRST ASSESSMENT OF ITEM INTER-RATER RELIABILITY

(Amended section)

For each item, the value of weighted Kappa (K_w) and its level of significance was calculated by MedCalc (Schoonjans, 2004). In addition, the value of K_w divided by the standard error of K_w was also calculated, because it is upon this (rather than the absolute value of K_w) that the level of significance depends.

The results, in descending order of significance, are listed below.

Table 9.1: Value and significance of gamma as measure of inter-rater agreement

No	Statement	K_w	K_w/se
31	The outcome is positive for the main character, but only at the expense of most others.	***.78	5.57
32	The outcome is negative for the main character.	***.72	5.54
37	The story as a whole seems to be pessimistic or negative.	***.76	5.07
3	There is much use of colour made in this picture.	***.81	4.76
2	The picture has obvious images of conflict, violence or death in it.	***.83	4.61
40	There are many mythological creatures or themes in the story.	***.67	4.47
35	The moral of the story is about persistence, brains or bravery achieving success.	***.57	4.38

No	Statement	K _w	K _w /se
7	Some characters in this story are superior, grandiose, of high rank, celebrated or admired.	***.76	4.22
30	The outcome is positive for the main character.	***.76	4.22
43	The settings, characters and events in this story mostly have magical or fantasy elements to them.	***.70	4.12
16	The main task is to capture, kill or eat some other creature or being.	***.86	4.10
28	Others are helpful because their presence means the main character is not alone.	***.69	4.06
17	Violent, physical means are used to overcome obstacles.	** .72	4.00
27	Other characters essentially complete the task for the main character.	** .76	4.00
46	The story as a whole seems to be an optimistic or positive one.	** .67	3.94
39	Magic powers, wishes or spells are important in this story.	** .55	3.93
1	Every part of this picture seems to be full of rich decoration and detail.	** .62	3.88
29	The outcome is a 'win-win' situation for main character and most others.	** .68	3.78
34	The moral is that we all have untapped potential in us.	** .51	3.64
38	Positive images of life, growth, health or production predominate.	** .51	3.64
22	The characters in this story show no awareness of one another's needs and give one another no consideration.	** .58	3.63
47	Things that harm others are important to the story – guns, swords, teeth, claws.	** .69	3.45
20	Obstacles are overcome with faith, persuasion, skill or problem-solving.	** .48	3.43
44	Themes of good and evil, right and wrong are important in this story.	** .61	3.39
10	There is an obvious bullying, dominating, violent character in this story.	** .59	3.28
4	In this story there is a rescuing, caring character.	** .49	3.27
25	There are many examples of interactions between characters in this story.	** .52	3.25
6	The main character has likeable, admirable qualities	* .54	3.18
21	The main character heroically overcomes stronger opponents.	* .57	3.17
41	The whole atmosphere of this story is barren, bleak and lonely.	* .47	3.13
19	Problems in this story tend to be about physical size and strength (or smallness/weakness.)	* .53	3.12
24	Others are mainly a threat towards the main character in the story.	* .50	2.94
15	Failure to complete the task would mean starvation or death for the main character.	* .47	2.94
23	Others are mostly seen as helpful, positive, friendly.	* .45	2.81
45	The content of the story is minimal, stark and brief.	* .44	2.75
12	Characters in this story seem to have belief and confidence in themselves.	* .35	2.69

No	Statement	K_w	K_w / se
5	The main character is specifically described as average, ordinary, normal.	.40	2.67
14	The main task is about a journey or escape from a hostile or unpleasant situation.	.48	2.67
18	Opposition makes the main character angry, frustrated...	.48	2.67
8	There is at least one important character who is larger than life – a god, superhero, film star, famous historical character or similar.	.30	2.50
11	At least one character in this story is sick, ill, poor or in need of rescue.	.35	2.50
13	The main task is of self-development, satisfying curiosity, enlightenment, achieving potential.	.35	2.50
36	The story teaches that we always need other people.	.30	2.31
42	Morbid themes of death, aggression, pain or decay predominate.	.33	2.20
26	Themes of abandonment and being left alone by others are prominent.	.22	2.20
9	There is at least one character who is a powerless, overwhelmed victim in this story.	.27	2.08
33	The story teaches that confronting fears and problems is necessary to happiness.	.12	1.71

* $p < .05$, ** $p < .01$, *** $p < .001$, all tests one-tailed.

It was decided that a satisfactory level of inter-rater agreement and item suitability had been reached where the value of weighted Kappa was significant at the level of $p < .05$, and that other items would be discarded. This resulted in the elimination of 11 items and the retention of 36 for the final phase where all stories in the study would be analysed and all raters would be asked to provide ratings.

It was recognised that there are two caveats with this approach. First, with 47 tests being undertaken, and a significance level of $p < .05$ being used, it was highly likely that two or more spurious positive results would be found without a Bonferroni correction being made. Second, it is not usual to use a significance level to define a cut-off; given that with a large enough sample most or all statements might achieve statistical significance.

However this was not meant to be the final, definitive selection of statements to be used in the full analysis. What was wanted was simply to exclude those statements that were poor candidates at this first stage. A more stringent test would be applied at the stage involving more raters. Secondly, the statistical significance depends upon both the value of Kappa and its standard error; where the former divided by the latter falls below a certain level (in this case 2.69) agreement ceases to be significant. It was because this threshold includes a term for the standard error as well as the absolute value of weighted Kappa that it was used at this stage.

10.0: RATER ANALYSIS OF STORIES: RESULTS

10.1 STRUCTURE OF CHAPTER

This chapter reports the results of the analysis of the ratings provided by the panel of raters. 65 stories were rated, of which 40 were rated by at least two raters, allowing inter-rater reliability of statements about the stories to be assessed. Those statements that could be rated reliably were then assembled into scales by making comparisons with the concurrent measures taken, and the scales assessed for inter-rater and test-retest reliability as well as concurrent validity by comparison with other clinical measures.

10.2 INTER-RATER RELIABILITY OF INDIVIDUAL ITEMS

Rather than having two raters each rating 40 stories, there were 24 raters each rating up to 8 stories. This was because each story took approximately 30 minutes to rate, and it was not felt practical to burden volunteer raters with more than four hours of rating work each. At this stage a total of 48 paired ratings of single stories (one rater versus another) were available for comparison.

The analysis of data provided by the blind raters started with an estimation of the inter-rater reliability of the individual items on the response form. Using the weighted Kappa statistic, inter-rater item reliability was tested on the batches of eight stories by two raters and only about half of the items had acceptable reliability (weighted Kappa significant at $p < .05$). This was despite all these items having been demonstrated to have adequate inter-rater reliability at the pilot stage when two raters rated a larger number of stories. With only eight

stories, each variable had only eight data points and it was feared that significant trends might be being obscured by the small numbers.

It was decided to treat the data from the 24 raters as if they had indeed been drawn from just 2, thus analysing one set of 40 stories rather than several smaller batches. It was recognised that this introduced another source of variation into the ratings being analysed. Weighted Kappa assumes that all the ratings in the rows and columns of the contingency table are being produced by one rater, not several. Any variability between ratings is therefore due to inconsistency on the part of the individual rater or variation between stories. Amalgamating 16 raters into two in this way introduces a possible inter-rater variability that will potentially reduce the level of agreement achieved.

However this would make weighted Kappa less sensitive to agreement, not over-sensitive. Type II errors would be more likely and Type I errors less likely. Any significant agreement discovered would be more likely to be reliable, not less. A Bonferroni correction (Bland & Altman, 1995) was not made because these statements had all been tested once before and found to be significantly reliable; therefore it was to be expected that a large number of significant associations would be found. A Bonferroni correction would have set the threshold for significance unrealistically high and introduced an unknown number of Type II errors (Perneger, 1998). As a compromise, a more stringent threshold for significance was set at $p < .01$. With 36 separate tests being conducted, this threshold made it unlikely that more than one spurious result (if that) would be obtained. In addition, only those significant results where the

absolute value of weighted Kappa was greater than .4 were taken as showing significant agreement.

Table 10.1: Inter-rater reliability of statements relating to stories and pictures

No	Statement	Kw
16	The main task is to capture, kill or eat some other creature or being.	***.77
30	The outcome is positive for the main character.	***.65
32	The outcome is negative for the main character.	***.62
3	There is much use of colour made in this picture.	***.61
17	Violent, physical means are used to overcome obstacles.	***.55
29	The outcome is a 'win-win' situation for main character and most others.	***.54
15	Failure to complete the task would mean starvation or death for the main character.	***.54
27	Other characters essentially complete the task for the main character.	***.52
46	The story as a whole seems to be an optimistic or positive one.	***.47
47	Things that harm others are important to the story – guns, swords, teeth, claws.	***.47
23	Others are mostly seen as helpful, positive, friendly.	***.46
22	The characters in this story show no awareness of one another's needs and give one another no consideration.	***.44
39	Magic powers, wishes or spells are important in this story.	** .44
2	The picture has obvious images of conflict, violence or death in it.	** .42
10	There is an obvious bullying, dominating, violent character in this story.	** .39
1	Every part of this picture seems to be full of rich decoration and detail.	** .38
28	Others are helpful because their presence means the main character is not alone.	** .37
37	The story as a whole seems to be pessimistic or negative.	** .34
6	The main character has likeable, admirable qualities	** .33
40	There are many mythological creatures or themes in the story.	*.30
41	The whole atmosphere of this story is barren, bleak and lonely.	*.29
21	The main character heroically overcomes stronger opponents.	*.26
44	Themes of good and evil, right and wrong are important in this story.	*.26
25	There are many examples of interactions between characters in this story.	*.25
31	The outcome is positive for the main character, but only at the expense of most others.	*.25
4	In this story there is a rescuing, caring character.	*.25
43	The settings, characters and events in this story mostly have magical or fantasy elements to them.	*.25
34	The moral is that we all have untapped potential in us.	*.24

No	Statement	Kw
12	Characters in this story seem to have belief and confidence in themselves.	*.22
24	Others are mainly a threat towards the main character in the story.	.22
38	Positive images of life, growth, health or production predominate.	*.21
45	The content of the story is minimal, stark and brief.	*.21
35	The moral of the story is about persistence, brains or bravery achieving success.	.19
7	Some characters in this story are superior, grandiose, of high rank, celebrated or admired.	.19
20	Obstacles are overcome with faith, persuasion, skill or problem-solving.	.16
19	Problems in this story tend to be about physical size and strength (or smallness/weakness.)	.09

* $p < .05$, ** $p < .01$, *** $p < .001$, all tests one-tailed.

Of the 36 items, 14 were retained as having a value of weighted Kappa $>.4$ and a significance of $p < .01$. The other 22 items were discarded.

Stories varied in having between one and four raters, and the average score for all raters was calculated for each of the remaining 14 items.

10.3 CONSTRUCTION OF SCALES

10.3.1 SCALE ANALYSIS OF REMAINING ITEMS

It had been hoped to use factor analysis to identify variables that grouped together, but there were insufficient rated stories (65 in total) for this to be a valid operation with 14 candidate variables. Nevertheless some systematic, replicable method of grouping variables was required and it was desirable to not yet use the concurrent data (CORE, SCID-II etc.) This was because to select

variables for their correlation with, say, the CORE-Risk score would mean that any scale thus assembled would have an artificial and unconvincingly high correlation with the CORE scores.

The method used was initially to inspect the mutual correlations of the 14 remaining variables. When the significant correlations were inspected it was apparent that there was a consistent pattern, with eight variables correlating positively with one another, and negatively with six others. The mean scores on each variable were in the range 0-4, so the inverse of the negatively correlated scores was derived by subtracting the variable score from 4. This gave 14 variables that were all positively correlated with one another.

The reason for making this transformation was that this made possible a scale reliability analysis, using all 14 variables as the initial candidates. The data were analysed using SPSS and variables removed from the list successively in order to maximise the value of Cronbach's alpha for the developing scale. In the first pass, eight variables were removed before alpha was maximised, leaving the following six items to comprise a scale with a Cronbach's alpha of .95:

Table 10.2: Components of first scale identified

Correlation with scale	No	Statement
+	22	The characters in this story show no awareness of one another's needs and give one another no consideration.
+	32	The outcome is negative for the main character.
-	23	Others are mostly seen as helpful, positive, friendly.
-	29	The outcome is a 'win-win' situation for main character and most others.
-	30	The outcome is positive for the main character.
-	46	The story as a whole seems to be an optimistic or positive one.

Statements 32 and 30 were so diametrically opposed as to make one of them redundant – in fact the correlation between these two statements was so high as to be almost perfect (Spearman's $\rho = -.92$, $n = 65$, $p < .001$). Statement 32 was removed from the scale as its value of weighted Kappa for inter-rater agreement was marginally lower.

This left eight statements unallocated, which were again put through the iterative process of scale analysis to maximise Cronbach's alpha. This time four statements remained when alpha was maximised, to comprise a scale with a Cronbach's alpha of .82. The four items (all contributing positively to the scale) are described in Table 10.3 below.

Table 10.3: Components of second scale identified

Correlation with scale	No	Statement
+	2	The picture has obvious images of conflict, violence or death in it.
+	16	The main task is to capture, kill or eat some other creature or being.
+	17	Violent, physical means are used to overcome obstacles.
+	47	Things that harm others are important to the story – guns, swords, teeth, claws.

After inspection of their components, the first scale was given the title 'Negativity scale' and the second 'Violence scale'. The next step was to ascertain the reliability of the scales as a whole, as well as their individual elements.

10.3.2 RELIABILITY OF SCALES

For each scale, a value was calculated by adding together the value of all items with a positive contribution and subtracting the value of all items with a negative contribution. The value of the two scales was calculated and recorded for each

of the 65 stories. The same values were also calculated for the ratings from individual raters, to compare inter-rater reliability on these scale scores.

Intra-class correlation was used to calculate agreement between the factor scores given by first and second raters. The results were as follows:

Table 10.5: Inter-rater reliability of scale scores

Scale	Label	ICC	F	df
1	Negativity	***.85	12.16	54
2	Violence	***.76	7.33	54

*** $p < .001$, all tests one-tailed.

Both scales appeared to have strong inter-rater reliability. A scatter plot was drawn for both scales showing ratings given by each rater on a separate axis. There appeared to be no distortion due to outliers or curvilinear relationships.

Where there were patients who had completed two stories, it was possible to estimate the test-retest stability of the two scales. Results were as follows:

Table 10.6: Test-retest reliability of scale scores

Scale	Label	ICC	F	df
1	Negativity	*.49	2.95	16
2	Violence	.31	1.90	16

* $p < .05$, all tests one-tailed.

Of the two scales identified, only the first (negativity) was adequate both on inter-rater and test-retest reliability. Only this factor was investigated further to see whether it was a valid predictor of any of the concurrent clinical variables measured.

10.3.3 CONCURRENT VALIDITY OF NEGATIVITY SCALE

The negativity score was compared with the other concurrent data collected via the IIP, CORE and SCID. It correlated positively with the summary scores for each of these measures. It also correlated positively with the rater estimation of mental health, inferred after reading each story:

Table 10.7: Correlation of negativity score with concurrent measures

Measure	Pearson's ρ	n
IIP mean score	*.51	22
CORE total score	*.47	22
SCID total criteria met	** .56	22
Rater inference of mental health	***-.66	64

** $p < .01$, * $p < .05$, all tests one-tailed.

Since the negativity score seemed to be most strongly associated with the SCID among the three validated concurrent measures, it was decided to investigate the three DSM-IV Axis II clusters to see whether any were more strongly associated with the negativity score than others. The results were as follows:

Table 10.8: Correlation of negativity score with SCID-II criteria met

Cluster	Spearman's r	p
A	.38	.08
B	** .67	.001
C	.11	.62

** $p < .01$, all tests one-tailed.

$n = 22$ in all cases

It appeared that the Negativity score was most closely correlated with the number of Cluster B criteria met, and had little or no correlation with Cluster A or C criteria. A scatterplot was drawn to inspect the relationship between the Negativity score and the number of Cluster B criteria met, and this confirmed a reasonably linear relationship undistorted by outliers (Figure 10.1 below.)

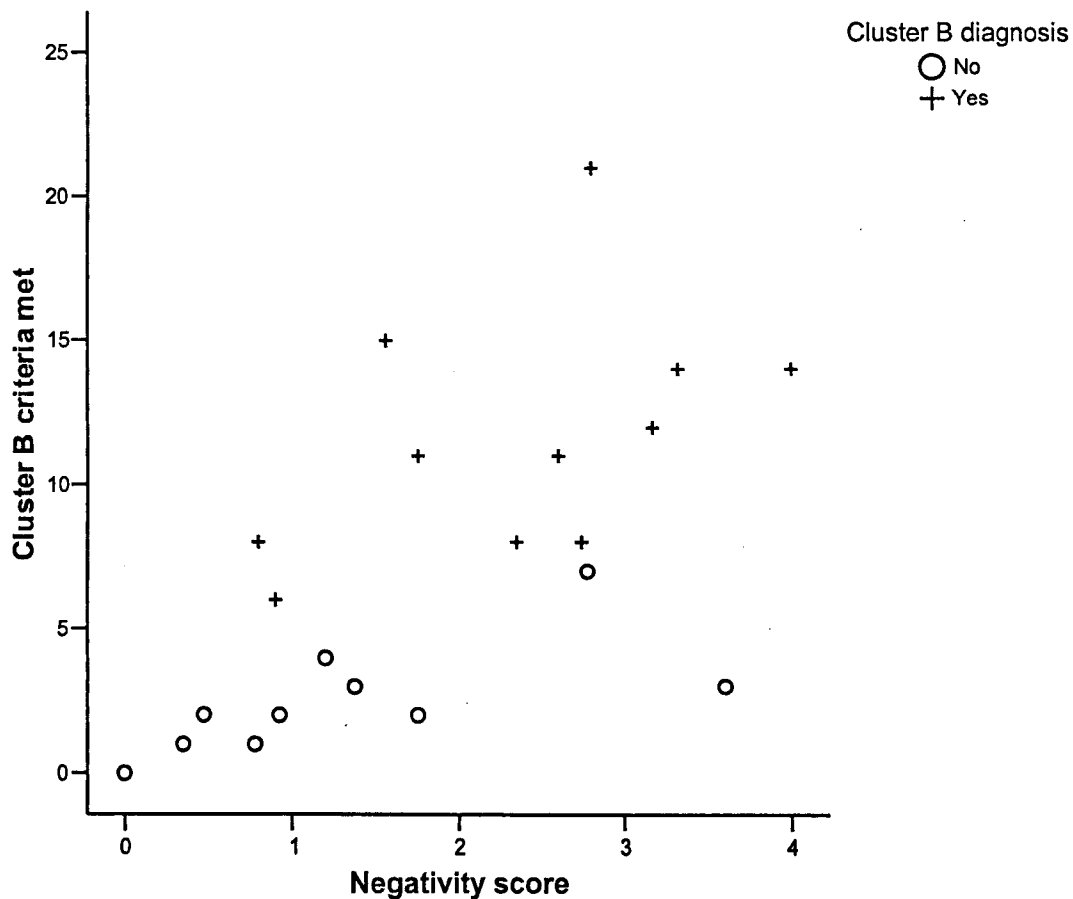


Figure 10.1: Scatterplot of Negativity score against Cluster B criteria

When the mean Negativity scores of those CMHT patients with and without a Cluster B diagnosis were compared, the difference approached, but did not reach, a significant level ($t = 1.95$, $df = 20$, $p = .066$). There were perhaps insufficient numbers involved (12 participants with a cluster B diagnosis and 10 without). If an assumption was made that none of the clinician participants had a cluster B diagnosis, and they were brought into the calculation, then the comparison became significant ($t = 3.65$, $df = 42$, $p < .05$). A boxplot was drawn (Figure 10.2 below) which suggested that the Negativity scores of CMHT patients without a cluster B diagnosis were much more similar to those of clinicians than those of patients with a cluster B diagnosis.

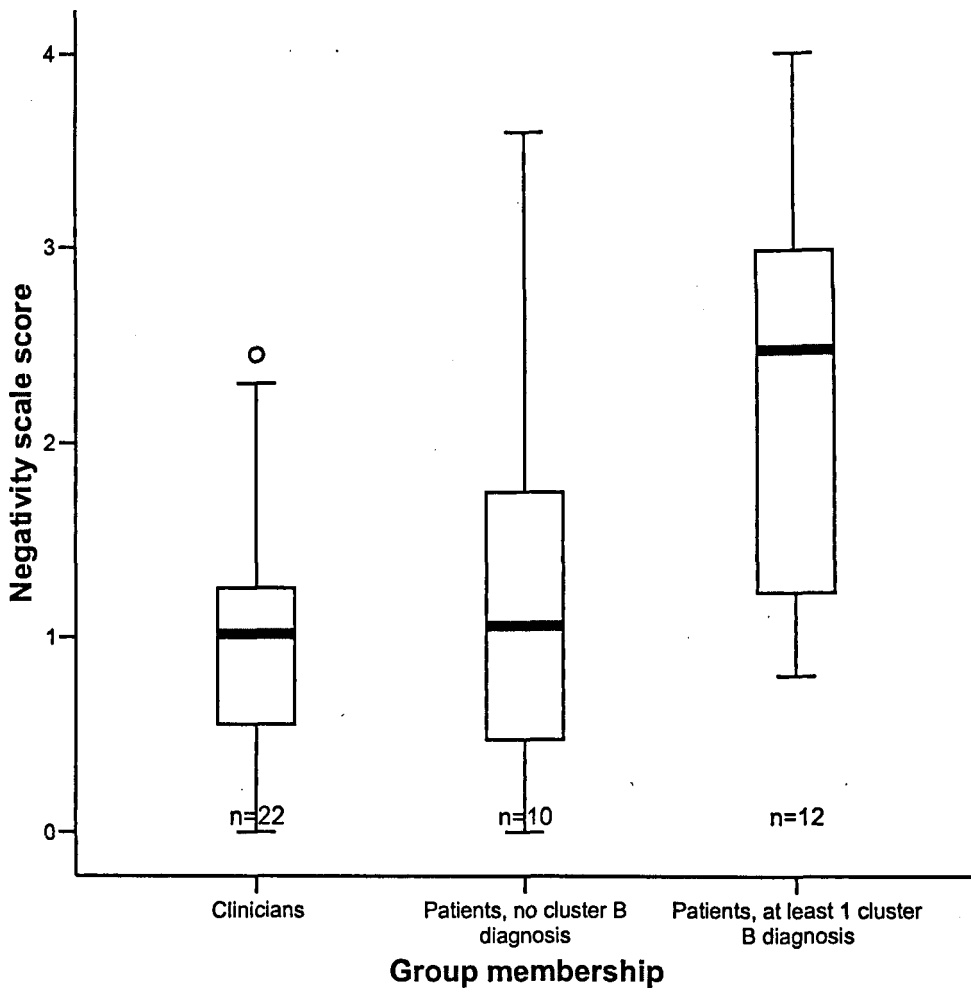


Figure 10.2: Boxplot of Negativity scores by group membership

It had already been noted that there was an imbalance of gender among those patients without a probable personality disorder (more male than female) and those with such a diagnosis (more female than male.) This left the possibility that the differences in negativity scores between the diagnostic groups were simply an artefact of the gender difference. To check this possibility, the correlation between negativity score and cluster B criteria was calculated, controlling for gender. There remained a significant correlation between the two (correlation = .45, $df = 19$, $p < .05$).

With the emergence of this strong theme of negativity, the possibility arose that this was linked to participants' depressive symptoms rather than their personality. To investigate whether the level of depression of the storyteller had an effect on the negativity scores of their stories it was necessary to use a proxy measure. No measure of depression had been taken as part of the measures given to participants, but a recent study has demonstrated that scores on the Beck Depression Inventory (BDI) can be inferred from patients' CORE scores (Leach, Lucock, Barkham, Noble, & Iveson, in preparation). Translation tables from this study were used to estimate the BDI score of all the patient participants.

The degree of depression was positively correlated with the negativity score (correlation = .45, $p < .05$, $n = 22$). Patient participants with a diagnosis of Borderline PD were scoring significantly higher on the inferred BDI score than others ($t = 4.02$, $p < .01$, $df = 21$) and it appeared possible that the negativity score might just be linked with depression and not personality disorder after all.

In order to gauge the relative importance of depression and personality disorder to the final negativity scores, an analysis of variance was carried out with the negativity score as the dependent variable. The independent variables were the inferred BDI score and number of SCID-II Cluster B criteria met; variables were entered stepwise in order that the variable with the greatest contribution to the variance should be entered first. The results of this analysis were that the contribution of the number of SCID-II BPD criteria was not significant (coefficient $\beta = .18$, $t = 0.62$, $p = .55$). Once this was removed from the analysis negativity score remained significant (coefficient $\beta = .45$, $t = 2.22$, $p < .05$).

To check this association, a partial correlation calculation was carried out to check the correlation of the Negativity score and the inferred BDI score while controlling for the number of cluster B criteria met. It appeared that when the number of cluster B criteria was controlled for, there was no significant correlation between the Negativity score and the inferred BDI score (correlation = .25, $df = 19$, $p = .14$).

It was not therefore possible to say with confidence whether the Negativity score was a reflection of the patient group's cluster B personality status or their degree of depressive symptoms.

10.4 STORIES FROM PATIENTS WITH OTHER PD DIAGNOSES

Although the main thrust of the study was on comparisons between those with and without a diagnosis of BPD, other Axis-II diagnoses had been identified. Further *t*-tests were undertaken on groups representing patients with and without diagnoses of Schizoid PD and Avoidant PD. Because of the strong comorbidity of the BPD diagnosis with diagnoses of Passive/Aggressive PD, Antisocial PD, Paranoid PD and Depressive PD in this sample these personality disorders were not investigated in this way. There were too few patients with diagnoses of Histrionic PD, Dependent PD, Narcissistic PD, Schizotypal PD and Obsessive/Compulsive PD for such investigations to be meaningful.

10.4.1 SCHIZOID PERSONALITY DISORDER

There were six stories from patients with a diagnosis of Schizoid PD and 34 from other patients. These were produced by four and 19 individuals respectively. When a *t*-test was performed on statements about stories produced by the two groups, seven statements were significantly different. However four of these had already been identified and used to make up the negativity scale that had been strongly identified with a cluster B diagnosis. The other three statements were taken together and summed into a single Schizoid scale score. The statements were:

- The main character has likeable, admirable qualities -
 - There is an obvious bullying, dominant, violent character in this story +
 - The moral is that we all have untapped potential in us -
- + indicates statement associated positively with diagnosis of Schizoid PD
- indicates statement associated negatively with diagnosis of Schizoid PD

When summed into a single score this scale proved to discriminate well between patients with and without a diagnosis of Schizoid PD ($t = -2.97$, $df = 36$, $p < .01$). The optimum cut-off for this score (established by drawing the ROC curve) gave the following contingency table:

Table 10.8: Predicted and observed diagnosis of Schizoid Personality Disorder

	Predicted positives (from Schizoid score)	Predicted negatives (from Schizoid score)
Observed positives (from SCID-II)	5	1
Observed negatives (from SCID-II)	6	26

31 out of 38 (82%) of stories would be correctly allocated, although this cut-off does produce more false positives than true positives.

10.4.2 AVOIDANT PERSONALITY DISORDER

The same procedure described above was then carried out on stories from those with an Avoidant PD ($n = 8$) and those without ($n = 30$). This time there were just two statements that distinguished the groups:

Failure to complete the task would mean starvation or death for the main character	+
Magic powers, wishes or spells are important in this story	-

+ indicates statement associated positively with diagnosis of Avoidant PD
- indicates statement associated negatively with diagnosis of Avoidant PD

When summed into a single score these two items discriminate well between patients with and without a diagnosis of Avoidant PD ($t = -3.87$, $df = 36$, $p < .001$). The optimum cuto

10.5 CORE CONFLICTUAL RELATIONS THEME (CCRT) RATINGS

As well as rating the 47 statements, raters were also asked to use the Core Conflictual Relations Theme (CCRT) categories to identify three themes in each story; a wish for the main character, the response of others and the main character's reaction to others' response. Raters chose from a list of eight possible categories in each of the three themes. Inter-rater agreement on assignment to the eight categories was assessed by calculating the Kappa statistic (not Gamma because the data were categorical, not ordinal.) The results were as follows:

Table 10.10: Inter-rater agreement within CCRT themes using 8x8 contingency tables

Theme	<i>Kappa</i>	<i>T</i>
Wish of main character	***.26	4.03
Response of others	***.37	5.61
Reaction of main character	***.22	3.62

*** $p < .001$, one-tailed.

These results were statistically significant across the body of 58 stories where at least two ratings were available. However the values of Kappa were not high and the statistical significance simply indicates that there were fewer disagreements than might have been expected by chance; there were in fact still many more disagreements between raters than agreements.

To derive a more clinically useful result, the eight possible CCRT categories within each theme were collapsed into two; a positive and a negative response. For example, the eight possible CCRT categories in the "Response of others" theme were collapsed thus:

Table 10.11: Collapsing the eight CCRT “Response of others” categories into two ratings (Positive and Negative)

Category	Rating
Others are strong, independent and happy	Pos
Others are strict and controlling	Neg
Others get upset. They are hurt, anxious, dependent, angry or out of control	Neg
Others are bad and cannot be trusted	Neg
Others reject and oppose the main character; they do not trust or understand the main character, they are disrespectful, unhelpful or hurtful.	Neg
Others are helpful and co-operative towards the main character	Pos
Others like the main character, love them, give them independence.	Pos
Others are understanding, open and accepting towards the main character	Pos

Once this collapsing from eight categories to two had occurred, the inter-rater agreements were as follows:

Table 10.12: Inter-rater agreement within CCRT themes after collapsing into 2x2 contingency tables

Theme	<i>Kappa</i>	<i>T</i>
Wish of main character	.24	1.85
Response of others	***.54	4.20
Reaction of main character	***.58	4.43

****p* < .001, one-tailed.

The reason for the decline in significance of the first category, “Wish of main character”, was that very few of the wishes were coded as negative, as shown in the contingency table for agreement below.

Table 10.13: Collapsed 2x2 contingency table For CCRT theme “Wish of main character”

Second rater	First rater		Total
	Positive	Negative	
Positive	44	6	50
Negative	5	3	8
Total	49	9	58

The wish of the main character was seen as positive most of the time, but where it was seen as negative by one rater there was very little agreement (3 out of 14 cases) with the other rater. Thus although the percentage agreement was higher than in the 8x8 contingency table, the significance was lower.

Because of the lack of inter-rater agreement in this first theme, it was not investigated further. The other two themes were investigated for test-retest stability in the following way. Where several raters had rated a single story, their ratings were amalgamated as a majority; for example if four raters coded the response of others as positive, and one rater coded it as negative, that story was coded as having a positive response of others. If raters tied in their ratings the story was excluded from the analysis at that point. Once stories had been recoded, those from authors who had produced two stories were compared to assess test-retest stability. The results shown in the table below suggest that only the theme of "Reaction of main character to others" has acceptable test-retest reliability.

Table 10.14: Test-retest reliability of remaining CCRT themes

Theme	<i>Kappa</i>	<i>T</i>
Response of others	.26	0.97
Reaction of main character	*.69	2.49

* $p < .05$, one-tailed.

Thus it appeared that, after collapsing responses into two categories (positive and negative), the CCRT ratings of theme "Reaction of the main character to others" had acceptable inter-rater and test-retest reliability. Stories from the three main groups under investigation were investigated to see how they were rated on this theme.

Table 10.15: Response of main character, by author group

Response of main character to others	Group			Total
	Clinician	No BPD	BPD	
Positive	14	13	2	29
Negative	3	7	12	22
Total	17	20	14	51

The responses of the three groups were significantly different ($\chi^2 = 15.39$, $df = 2$, $p < .001$), with the largest contribution to the value of chi-squared coming from the increased proportion of negative responses to others rated in stories from authors with a BPD diagnosis.

Stories where the response of the main character was rated positively were compared to those where the response was rated negatively. On every concurrent clinical measure, the mean mental health of the stories' authors was poorer among those stories rated negative. The measures assessed by *t*-test were the four CORE subscales, the CORE total, IIP mean, total number of SCID-II criteria met and total number of PDs diagnosed by SCID-II. The lowest level of significance was for the CORE Risk subscale ($t = -4.11$, $df = 24.99$, $p < .001$, equal variances not assumed.) All other significances exceeded this.

10.6 RATER INFERENCE OF GLOBAL MENTAL HEALTH

A further element of the rating form was for raters to make an estimation of the mental health of the author of each story, using a simple 1-10 scale from poorest imaginable to best imaginable. There was a significant, moderate inter-rater correlation between scores given by the two raters for each story ($r = .46$, $n = 56$, $p < .001$, one-tailed test). Where two stories had been completed on

separate occasions, there was a significant, moderate test-retest correlation ($r = .64, n = 17, p < .01$, one-tailed test).

But what, if anything, was this global and very general statement actually measuring? Having established that the rating seemed to have adequate test-retest and inter-rater reliability it was compared with the concurrent clinical data and the inferred story rating data gathered thus far. The rating of mental health correlated negatively with all the concurrent measures from the CORE, IIP and SCID-II at the level $p < .01$ or better. The strongest correlation was with the total number of SCID-II criteria met ($r = -.63, n = 38, p < .001$, one-tailed test). The weakest correlation was with the CORE Risk subscale ($r = -.42, n = 38, p < .01$, one-tailed test).

The three main groups being investigated (clinicians, patients without a BPD diagnosis, patients with such a diagnosis) were compared to see how they were rated on this inferred global measure of mental health. In the boxplot below, it did appear that there were differences between the groups.

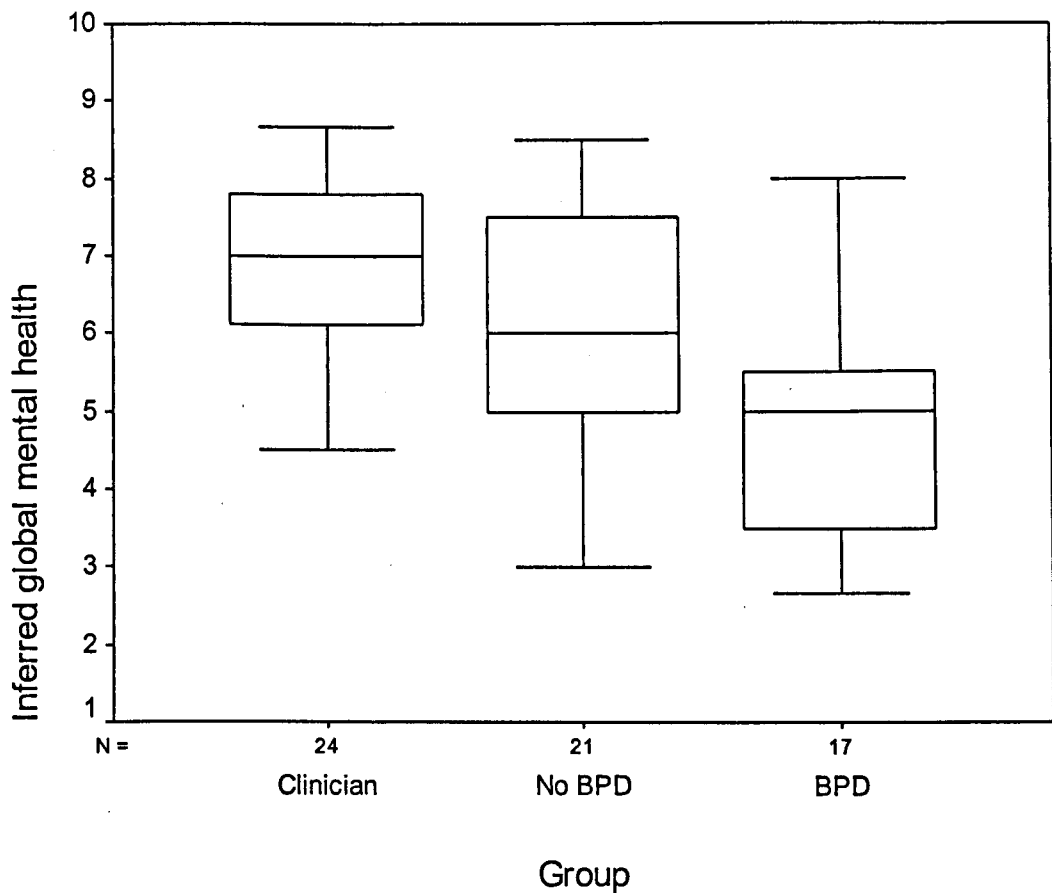


Figure 10.4: Inferred rating of global mental health by group membership of story author

As might be expected, there was reluctance on the part of raters to use extreme scores. Out of 176 separate ratings, only 28 were scored 1, 2, 9 or 10. As a result ratings were concentrated in the middle of the range between 3-8.

Nevertheless, the three groups were rated significantly differently. Stories from patients with a BPD diagnosis were given a significantly lower rating of mental health than those from the other two groups combined ($t = -3.74$, $df = 28.30$, $p < .01$, equal variances not assumed.) Stories from the two patient groups were also rated significantly differently ($t = 2.16$, $df = 36$, $p < .05$). The ratings given to stories from clinicians and patients with no BPD diagnosis were not significantly different however ($t = 1.86$, $df = 32.48$, $p > .05$, equal variances not assumed.)

10.7 RATER INFERENCE OF AUTHOR GENDER

The final element of the rating form was for raters to guess the gender of the author. The results of these estimations were:

Table 10.16: Inter-rater agreement on estimation of author gender

Second rater's guess	First rater's guess		Total
	Male	Female	
Male	21	10	31
Female	10	15	25
Total	31	25	56

Raters agreed with one another in 36 cases and disagreed in 20 cases. This was a modest level of agreement with one another. ($Kappa = .28$, $T = 2.08$, $n = 56$, $p < .05$) More importantly, their degree of agreement with the actual gender of the author could be measured in a similar way.

Table 10.17: Accuracy of rater estimation of author gender

Actual gender	Inferred gender		Total
	Male	Female	
Male	35	11	46
Female	57	62	119
Total	92	73	165

Altogether there were 165 ratings made. Of the 165 inferences about gender, 97 were correct and 68 incorrect. Raters correctly identified the gender of the author 58.7% of the time. With 41.3% of guesses being incorrect this is of little clinical utility, but this was nevertheless a statistically significant level of agreement ($Kappa = .22$, $T = 3.27$, $n = 165$, $p < .01$). Raters were correctly identifying the gender of the author significantly more often than would be

expected to occur by chance. The low value for Kappa however means that this statistical significance is of little practical significance.

It is interesting to note that when the author was male, this was correctly identified 35 times out of 46 (76.1%) of the time, whereas when the author was female this was correctly identified only 62 times out of 119 (52.1%). However, main characters in all stories were overwhelmingly male (both male and female tellers chose male characters much more frequently than female). This may have led raters to overestimate the number of male tellers; 92 of the 165 ratings (55.8%) inferred a male teller whereas of the 64 stories rated only 19 (29.7%) actually had a male author. Raters cannot therefore be said to be having much success at identifying gender from the story content alone.

10.8 ASSOCIATION OF RATER MEASURES WITH OTHER FEATURES OF STORY DATA

Finally, a number of characteristics of stories and tellers were compared to one another, using a chi-square test in the case of categorical data, a Spearman's correlation in the case of scale data and a t-test in the case of mixed (scale vs categorical) data. The characteristics tested were:

- Gender of teller (categorical)
- Gender of main character (categorical)
- Type of main character (categorical)
- Inferred mental health (scale)
- LIWC-B score (scale)
- Pessimism failure score (scale)
- Education of teller (categorical)

The results of these analyses were either non-significant or unsurprising, and revealed nothing that had not already been established by earlier analyses. For the record, these results are included at Appendix 5.

10.9 THE EFFECT OF RATER CHARACTERISTICS ON STORY RATINGS

10.9.1 ASSESSING BIAS IN RATINGS OF STORIES

It had been hoped to use the ratings of clinician stories to investigate any bias in the ratings given by those same clinicians. For example, did clinicians whose stories were rated as more pessimistic by other raters in turn rate other people's stories as more (or less) pessimistic than might be expected? Unfortunately, although stories from 24 clinicians were rated, only five of these clinicians themselves provided ratings of other people's stories. There were therefore very few points of comparison. For example, the inferred rating of global mental health made by a clinician over the eight stories they rated was compared with the inferred rating given to their story by other raters and was not significantly correlated ($r = -.20$, $n = 5$, $p = .75$) but with only five data points it is quite possible that a strong trend is being missed. The scatterplot is well spread, but only a few more data points along either diagonal would make it appear that a strong correlation either way was possible.

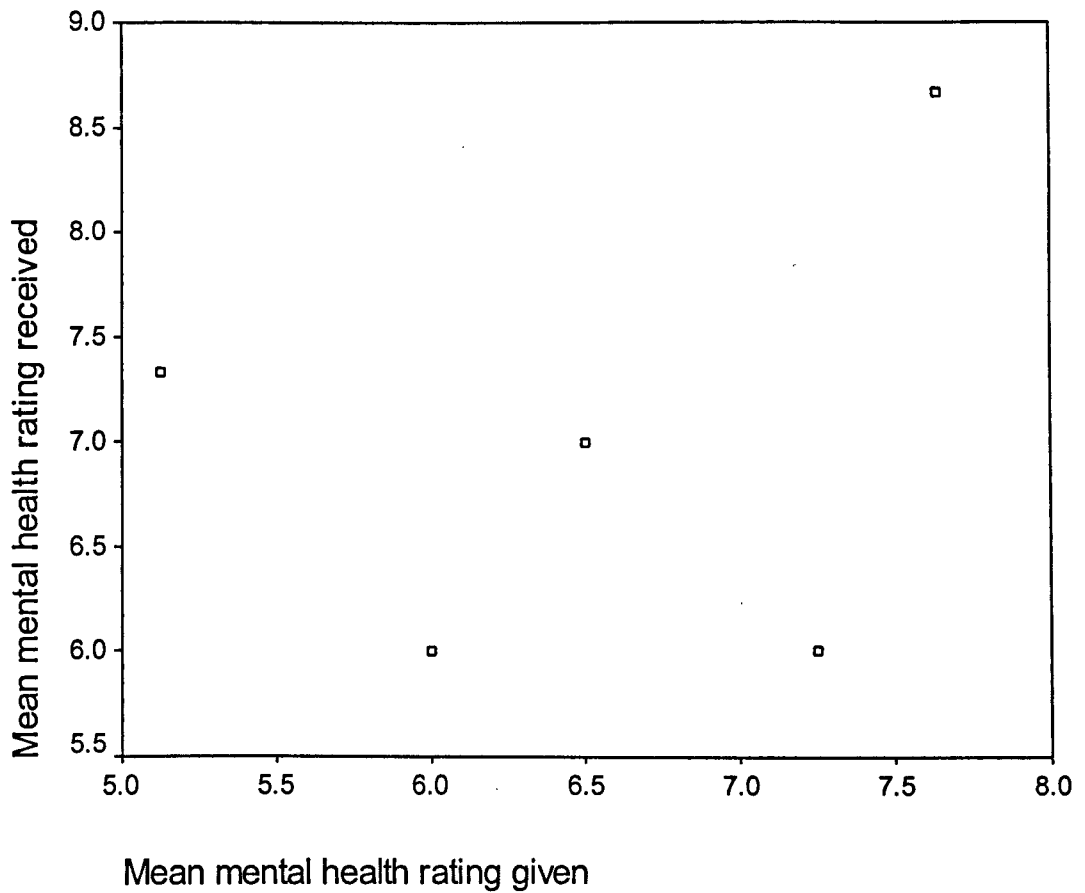


Figure 10.5: Mean mental health rating given and received by clinician raters

A similarly scattered distribution was found when the average Negativity scale score given by a rater was correlated with the average Negativity score given to their story by other raters ($r = .55$, $n = 5$, $p = .33$). Because of the unlikelihood of detecting any trend with such a small dataset, further attempts to assess rater bias in this way were reluctantly abandoned.

10.9.2 INFLUENCE OF RATER EXPERIENCE

Raters had been asked to indicate how familiar they were with the 6-Part Story Method by reporting how many stories they had heard. Four categories were reported: the numbers of raters in each category are given below:

Table 10.18: Number of stories previously experienced by raters

Number of stories experienced	Number of raters
1-4	10
5-10	9
11-20	3
More than 20	3

Thirty-one of the stories in the study had been rated by at least four raters; two less experienced (1-4 stories) and two more experienced (5 or more stories). It was therefore possible to compare the inter-rater reliability of two more experienced raters against two less experienced raters. The ratings given for global mental health and the six-item negativity scale were compared.

On the single 0-10 rating of global mental health, the more experienced raters had a strong inter-rater agreement ($r = .88, n = 7, p = .009$), noticeably better than the moderate inter-rater agreement among the less experienced raters ($r = .50, n = 30, p = .005$). The fit with the concurrent clinical data was tested by inspecting the correlation of the rating of global mental health given with the total number of SCID-II criteria met. For more experienced raters the correlation was strong ($r = .77, n = 21, p < .001$) but it was only slightly less strong for the less experienced raters ($r = .67, n = 21, p < .001$).

On the Negativity scale the more experienced raters also agreed strongly ($r = .83, n = 31, p < .001$), this time only slightly better than the less experienced raters ($r = .75, n = 31, p < .001$).

11.0 RATER ANALYSIS OF STORIES: DISCUSSION

"A clinician's delight and a statistician's nightmare" (Vane, 1981: p319)

11.1 INTER-RATER AGREEMENT ON STATEMENTS

It was interesting to note that the three statements which asked for ratings of the visual images attracted high levels of inter-rater agreement. They were ranked 4th, 5th and 16th out of 47 in terms of the value of weighted Kappa ascribed to them. However they did not subsequently make their way into any of the potential scales assembled to differentiate between stories from different groups of authors. Ratings of these visual elements were highly reliable therefore, but had no discriminative or predictive utility.

All 36 statements used by raters in the final study had previously been shown to have good test-retest reliability when two experienced raters rated a pool of 20 stories from the dataset. However only 14 of these statements could be rated by the larger pool of raters in the final study with acceptable inter-rater reliability. This difference could have been because of the relative experience of the raters; experienced raters presumably have a pool of all the other stories they have heard against which to set a rating of the current story. Some of the more inexperienced raters in the final study may only have heard two other stories before being asked to rate a story in this study, and what appears to them for example as an extreme story may be rated more moderately by raters with a wider experience.

In addition, there is the variability introduced by pooling the ratings from all 24 raters as if they had been just two. This introduced an extra source of variation which may have reduced the true level of agreement. However, if the effect of these two influences was to make the measure of agreement a more

conservative one this was felt to be acceptable. The 14 remaining statements about the stories had been assessed as having acceptable inter-rater reliability by both experienced and naive raters.

11.2 SCALE ASSEMBLY FROM STATEMENTS

The scale assembly process produced two possible scales, with component statements relating to the presence of negative, pessimistic themes and of violent imagery or events in the story. Both sets of statements had good face validity as a group, and appeared likely to have some kind of association with personality disorder. Inter-rater reliability for the two scores calculated from these scales was good, as might be expected when the component statement scores had already been selected for inter-rater reliability.

However only the first and longer scale, relating to the degree of negativity expressed in the story, had adequate test-retest stability. The other scale, although it could be scored reliably, did not seem to be stable from one occasion to another. This is not to say that the identification of violent images in the stories might not have some clinical significance; but the assumption was made that personality is a relatively stable substrate and that measures taken one month apart would show significant similarities. The discarded scale could have been a reflection of shorter term concerns of the tellers such as recent life events or fluctuations in mood, but as no concurrent measures had been taken against which this could be tested this were not further considered.

The final five items in the negativity scale were these:

- The characters in this story show no awareness of one another's needs and give one another no consideration. +
 - Others are mostly seen as helpful, positive, friendly. -
 - The outcome is a 'win-win' situation for main character and most others. -
 - The outcome is positive for the main character. -
 - The story as a whole seems to be an optimistic or positive one. -
- [+ indicates positive correlation with scale, - indicates negative correlation.]

Stories rated with high levels of negativity might have been expected from the mainstream CMHT group of patients, the majority of whom were suffering from depressive illnesses. It is somewhat surprising therefore that the negativity scale scores of stories from this group were indistinguishable from those produced by clinicians (see Figure 10.2). This does suggest that the negativity score is genuinely tapping a trait rather than a state.

Several of the statements that had been predicted to differentiate between BPD and non-BPD authors did not in fact do so; these included items such as:

- 'At least one character in this story is sick, ill, poor or in need of rescue.'
- 'The characters in this story show no awareness of one another's needs and give one another no consideration.'
- 'Themes of abandonment and being left alone by others are prominent.'

It is interesting to note that these were the kind of statements contained in the most obvious factor from the initial two-story pilot; this factor has not been replicated in the main study. In the pilot, obviously borderline features were strongly identified by the 19 raters as differentiating the two stories they read, whereas the larger number of stories in the main study lack this strong borderline factor. The two stories for the pilot were deliberately chosen to exemplify the differences between a story from a patient with BPD and a patient without. It is possible that when faced with two stories, the 19 pilot raters could

make strongly dichotomous choices between them. When a larger number of raters were faced with a wider range of stories from people across the spectrum of personality disorder, such differences were not so obvious.

The fact that the borderline features listed above seem to be just as common among stories from both groups of authors is a useful caution to practitioners who are already using the 6PSM as an assessment tool.

The five statements that do differentiate well, however, may be the basis for the validation of the 6PSM. These five statements are mutually consistent but do not immediately seem to define the borderline experience. The stories from patients with BPD do not consistently show the predicted elements such as dependency on another's idealised rescuing or clinging in the face of threatened abandonment. Perhaps the consistent theme coming from these stories is a depiction of what has been called abandonment depression (Manfield, 1992). This is characterised as the awful, isolated state against which borderline defences are erected so strongly; it may be an easier place to describe through the extended metaphor of the 6-part story than through personal introspection and disclosure.

Two sample stories may illustrate this further. Story one was produced by a patient with a diagnosis of BPD who scored high on the concurrent measures (CORE total z-score +1.42, IIP mean z-score +1.31).

STORY ONE:

Once upon a time there was this bear in a dark room, he was scared and he wanted to get out, so he was looking, thinking how to get out but he couldn't get out. He's just in the dark, it's like a room, it could be anything and the only way out of it is to get the key, but the thing that stops him getting the key is the brick wall. So the only thing he has is a hammer and some dynamite, so he thinks right, I'll blow up the wall because he can't get the key because of the brick wall. To get out of the building. And so he uses the explosive to blow down the brick wall, which knocked him down, and the wall has broken down and he's ended up in a bad state but yet the room where the key is is dark.

Story two was from another CMHT patient with no Axis II disorder, scoring much lower on the other measures (CORE total z-score -1.65, IIP mean z-score -1.2).

STORY TWO:

Once upon a time there was a boy who was at school and he had to do his exams. So he has got to learn lots of things so that he can do his exams and it is quite difficult because he wants to go out with his friends and it is boring and there is loads of stuff to be learned and it is a lot of hard work. But he has got his computer and he can look on the internet and he has got a teacher who can explain everything and make it more interesting. So he has to sit and read lots and lots of stuff and try and remember it all and test himself, and then he has to go to his exam and he has to write it all down. And then he passes his exam so he is really

happy and he spends lots of time in bed and lots of time down the disco having fun.

The first story starts from a bad place and ends in a worse one; it could even be said that the main character's attempts to improve his situation constitute a form of self-harm. There are no others around to help and the problem-solving attempt and final outcome involve physical violence and hurt. The second story, while somewhat prosaic and unimaginative, has a resourceful main character who receives help from others and eventually succeeds. In interview, the moral drawn by the first author was "The grass isn't always greener on the other side" while the second author said "You don't get a sense of achievement unless you have put in a fair amount of work, which is sometimes boring."

What is interesting is that the level of negativity and prediction of failure demonstrated by the first story was not apparent in the immediate presentation of the author. This was a patient who also fulfilled the criteria for a diagnosis of Antisocial Personality Disorder and came across in interview as confident, self-assured and in control. It may be that the 6PSM allowed this patient to reveal a more complete self-image than a straightforward interview would usually allow.

11.3 PSYCHOMETRIC PROPERTIES OF NEGATIVITY SCALE

On the face of it, this scale has adequate test-retest and inter-rater reliability, and it may discriminate between stories from authors with and without a cluster B diagnosis. It is not being proposed that the 6PSM should be used as a stand-alone test to diagnose the presence of a personality disorder. Instead it is hoped that the 6PSM and any scales derived from it might sensitise clinicians to the possibility of a cluster B disorder being present. As such a number of false positives are not a problem, particularly when the number of false negatives is so low.

The scales from text analysis had been developed by a bootstrapping approach, first correlating items with the CORE-Risk subscale. It is not surprising that scales derived from correlations with clinical data should then themselves correlate with other clinical measures; what is more interesting is that a scale arrived at by observing naturally occurring patterns in the data should also correlate in this way.

However the biggest problem with the scales derived at from textual analysis and factor analysis of human ratings is this: the scales have been derived from one set of stories and then tested on the same set. There is a possibility that the scales derived may capitalise on random variations in the data and therefore that their ability to discriminate may be an overestimate. Ideally, these scales should now all be tested against a new sample of stories to see the true level of discrimination achieved. To address this problem a split-half approach was tried; randomly splitting the sample, developing scales along the lines described

on one half and then testing them on the other half. Unfortunately although the scales closely resembled those derived from the full sample, the numbers involved were too small to demonstrate statistical significance.

11.4 INFLUENCE OF AUTHOR GENDER

The difference in gender balance between the cluster B and non-cluster B patient groups was a cause for some concern, because it raised the possibility that any differences between the groups might be due to gender rather than diagnosis. However it has been demonstrated that author gender has little, if any, effect on the ratings given to stories. Given the imbalance in gender between the groups, it was encouraging that the negativity scores were shown to be independent of gender; indeed, had this not been shown it would have been difficult to assert that the negativity score was in any way associated with personality disorder status.

Perhaps this similarity in the way men's and women's stories were rated goes some way to explain why raters found it so hard to guess the author's gender correctly. They were left with the choice of main character as the principal clue in guessing author gender, but it is an unreliable one. The only feature of the story that seems to differ significantly between male and female authors is in the gender of the chosen main character. Men almost without exception choose a main character of the same gender, while women do so a minority of the time. In other words most women and nearly all men choose male main characters, with a minority of women bucking the trend and choosing a female character.

11.5 CHARACTERISTICS OF STORIES FROM PATIENTS WITH SCHIZOID PERSONALITY DISORDER

When these results are based on just six stories from four individuals with a diagnosis of Schizoid PD it would be unwise to place too much reliance on them. However they may be suggestive of further work with this group. The three statements that seemed to significantly distinguish stories from the Schizoid PD group from others were:

- The main character has likeable, admirable qualities -
 - There is an obvious bullying, dominant, violent character in this story +
 - The moral is that we all have untapped potential in us -
- + indicates statement associated positively with diagnosis of Schizoid PD
- indicates statement associated negatively with diagnosis of Schizoid PD

This makes good sense if the main character and their qualities are taken as a representation of the author. It suggests a denigratory self-image with no sense of personal agency, in a setting with dominant, oppressive others. This picture of the schizoid world is very close to that described by Manfield (1992)

When formed into a scale, these three items are over-inclusive, identifying more false positives than true positives. But the scale only produced one false negative in six stories, and this may be a pointer towards a useful screening approach. Schizoid PD has a much lower profile within the mental health and psychotherapy literature and by their very presentation Schizoid PD patients are likely to be much less visible to their key workers than those with Borderline PD. And yet the UK nationwide prevalence of Schizoid PD is marginally higher; eight cases per thousand for Schizoid PD compared to seven per thousand for BPD (Singleton *et al.*, 2003: p68). When a patient produces a story that scores above the Schizoid threshold calculated here, it may be useful to consider the

possibility of a Schizoid PD diagnosis by more rigorous means, such as a diagnostic interview.

11.6 CHARACTERISTICS OF STORIES FROM PATIENTS WITH AVOIDANT PERSONALITY DISORDER

There were just two statements that differentiated this group from others:

- Failure to complete the task would mean starvation or death for the main character +
 - Magic powers, wishes or spells are important in this story -
- + indicates statement associated positively with diagnosis of Avoidant PD
- indicates statement associated negatively with diagnosis of Avoidant PD

The first statement can perhaps best be seen as an expression of the way the Avoidant PD patient sees the world; as a hostile, threatening place. Interestingly though, the obvious avoidant strategy of withdrawing from the task would imply a bad outcome, so perhaps there is some acceptance of the fact that this hostile world does need to be engaged with. The second statement is harder to read; it may be that this is an expression of impotence and hopelessness, or it may be a more realistic acceptance that just wishing for a solution is not going to bring one about.

The scale derived from these two statements has better specificity than that for Schizoid PD, with only two false positives. However with such small numbers this again can only be seen as a tentative and preliminary result needing further investigation.

11.7 CORE CONFLICTUAL RELATIONS THEME (CCRT) RATINGS

It was disappointing but not unexpected that agreement on the CCRT categories was poor. Agreement was statistically significant when looking at the full set of stories together, but on a story-by-story basis agreement was too low to be of clinical use.

However it was possible to find one feature from the CCRT which did have good inter-rater and test-retest reliability, and which did seem offer the possibility of clinical significance. The rating of the main character's response to others in the story as positive or negative seemed to form a continuum from clinicians through non-BPD patients to BPD patients. As many as 12 stories by the latter group (from a total of 14) were rated as having a negative response of the main character towards others. This may be an avenue that would be fruitful to explore in the clinical setting with individual patients, particularly if the main character's response to others in the story is seen as a metaphor for the author's response to others in the real world.

11.8 RATER INFERENCE OF GLOBAL MENTAL HEALTH

The correlation between the global rating of mental health and the number of SCID-II criteria met was striking. It is interesting that while authors do not seem to betray their gender through the stories they tell, they do seem to communicate something about their global level of personality function. This however depends on the total number of SCID-II criteria being used as a kind of

crude dimensional measure of personality function, which is not of course what the categorically-based SCID-II was designed to do.

The fact that the global rating of mental health correlated more highly with the negativity score than the LIWC-B score derived from text analysis is also to be expected. It would be surprising if the trends from the LIWC, involving much less visible phenomena such as pronoun use, were to be picked up by such a global question.

It is not being suggested that this global rating should be used in everyday clinical practice; the more detailed rating questions offer more sensitivity and clinical usefulness. But from a research point of view it is striking that such a simple question can produce good triangulation with both blind ratings and diagnostic interviews.

11.9 COMPARISON OF MORE EXPERIENCED AND LESS EXPERIENCED RATERS

Most raters in the study had heard fewer than ten stories. For many the only experience of the 6PSM they had had was the one-day training they had attended and they had not used it themselves in clinical practice. By contrast a smaller number of raters used the method regularly and had heard 20 stories or more. Given this imbalance it was expected that the more experienced raters would make more reliable ratings, and this proved to be the case. The difference however was not as great as expected; even relatively inexperienced raters could demonstrate good agreement with one another and could discriminate well between stories from BPD and non-BPD authors.

The only measure on which the more experienced raters performed markedly better was in the 0-10 global rating of mental health. Their inter-rater reliability was only moderate, as might be expected. It is perhaps less difficult to make statements about the content of the story (such as those that contribute to the negativity score) than it is to make inferences from the story about the teller's mental health. In making the former judgement no past experience is being called on. In the latter case it can be imagined that having heard a number of stories from people with different levels of mental health will give a more secure basis for making a judgement.

It appears then that ratings based on the manifest content of the story are probably to be preferred to ratings that explicitly demand that an inference be made that goes beyond the story material. Less experienced raters are more

able to agree on ratings that stay with story content. More experienced raters however can make ratings that go beyond this to directly make valid inferences about the story's author.

12.0 QUANTITATIVE TEXTUAL ANALYSIS

METHOD AND RESULTS

12.1 INTRODUCTION

This chapter describes the computer-based analysis of the transcribed text of the 67 stories, and the comparison of the results of this with the concurrent data gathered at the SCID interview and from self-report questionnaires.

The text analysed for this section comprised the transcript of the story as first told uninterrupted by the subject, along with their responses to the elaborating questions about the story itself. Responses to the final questions about the storymaking process were not included in the analysis. Tapes were transcribed and questions from the clinicians excluded from the transcript before analysis by the various computer programs.

A single SPSS master file was constructed, with each story acting as one case. The output of the various text analysis programs was pasted into the master file, using the unique tape number to identify each story. Data from the concurrent measures (SCID, CORE and IIP) were also pasted into the master file, using the identifier for the teller of the story as the key variable.

12.2 INVESTIGATIONS OF TRANSCRIPT LENGTH

12.2.1 STABILITY

Where two stories came from the same author, the length of transcript (initial telling of story plus responses to questions) at time 1 and time 2 could be compared. The transcript lengths on the two occasions were highly positively

correlated ($r = .90, n = 18, p < .001$). This relationship is demonstrated in Figure 12.1:

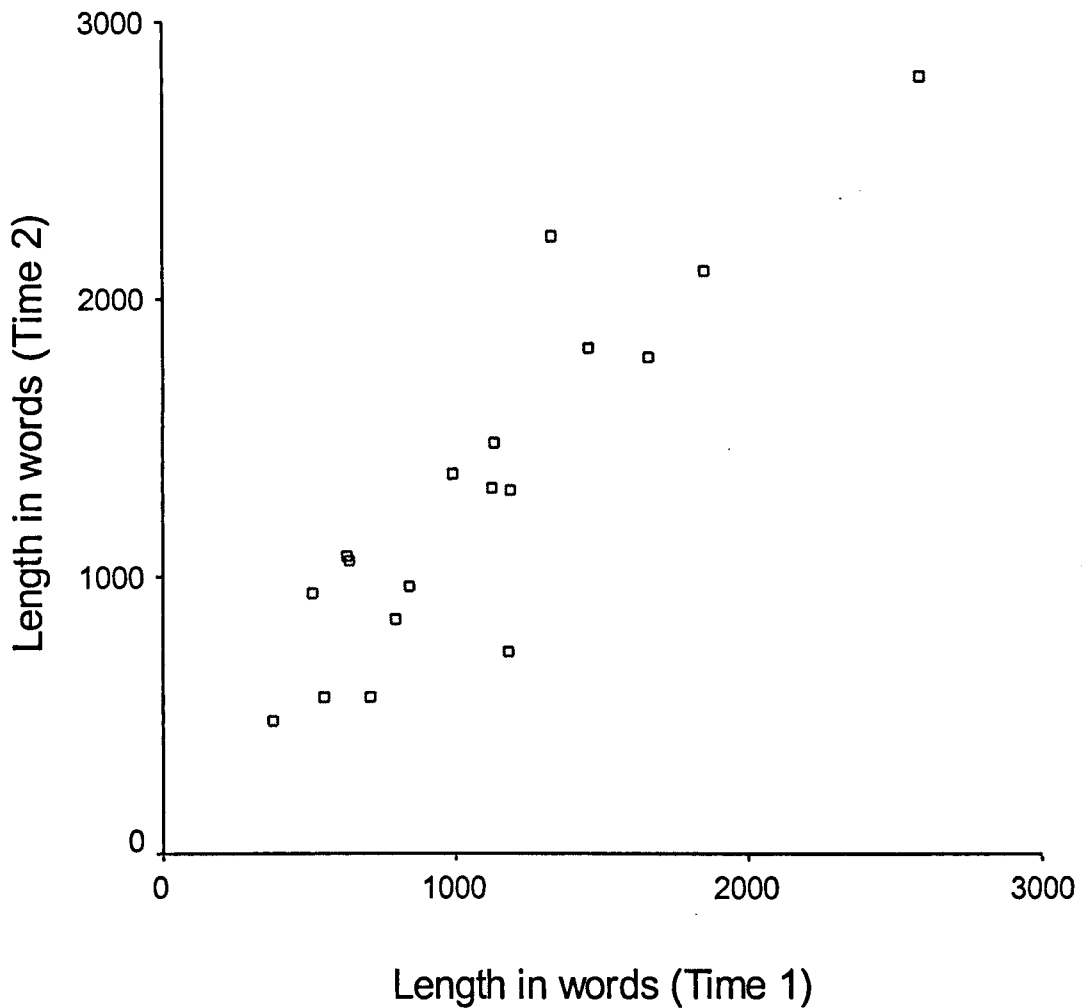


Figure 12.1: Scatter plot of lengths of first vs second story for the 18 patients who produced two stories. Each data point represents two stories from a single patient participant, recorded one month apart.

12.2.2 VARIATIONS IN STORY LENGTH

There appeared to be an inverse relationship between the length of a transcription and the complexity of personality disorder identified in the author. Patients were described as having a simple personality disorder if they had a PD diagnosis from one of the three Axis II clusters only. If they had diagnoses from two or all three Axis II clusters they were described as having a complex

personality disorder. The group of clinicians had not been given the SCID-II interview and no diagnosis of personality disorder is available for any of this group. However, their level of function was high enough to gain a qualification and continue in professional employment so it might be assumed that this group suffer few, if any, personality disorders. They are included in Figure 12.2 below for comparison with the three patient groups.

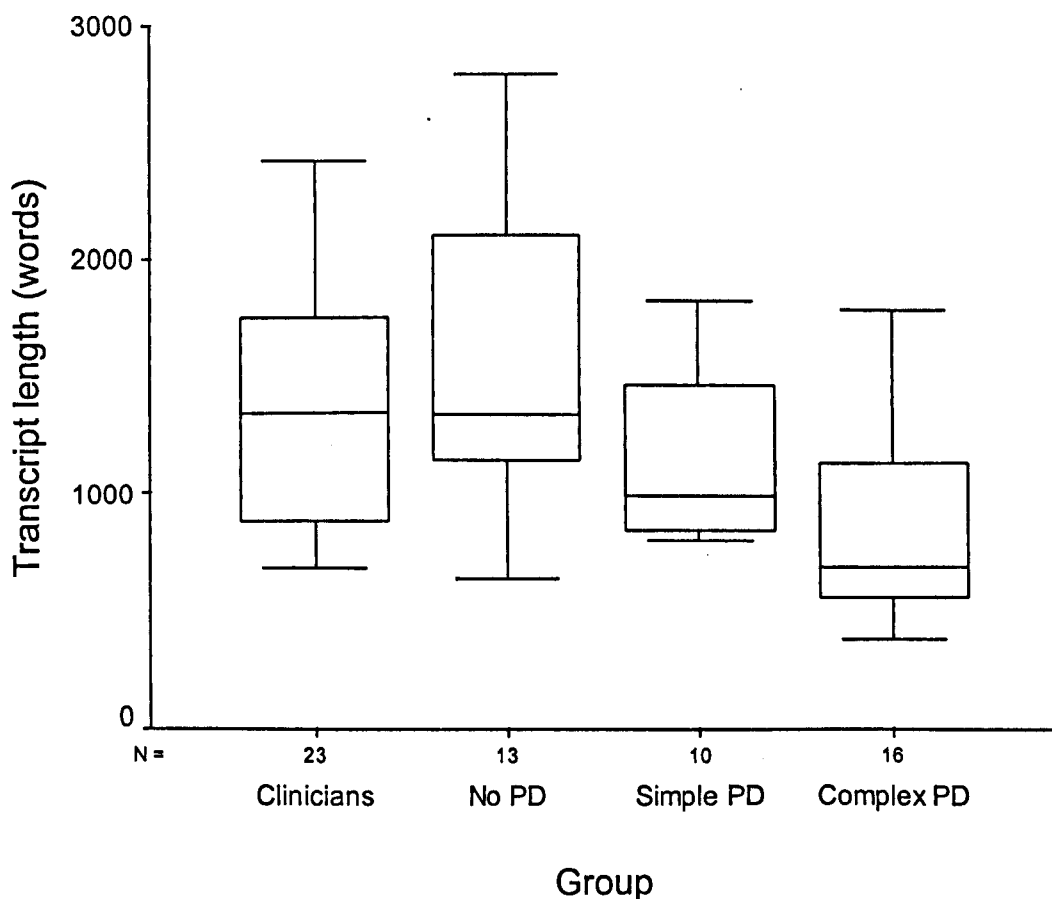


Figure 12.2: Boxplot of transcript length by complexity of personality disorders diagnosed. Central line represents median value, box represents inter-quartile range.

Thirteen stories came from patients with no personality disorder, and 26 from patients with at least one personality disorder diagnosed by the SCID-II. The mean word length of transcriptions from these two groups was compared and

found to be significantly different (two tailed t-test, $t = 3.09$, $df = 17.37$, $p < .01$, equal variances not assumed.)

The 23 stories from clinicians were also significantly longer than those from patients with at least one PD diagnosis (two-tailed t-test, $t = 2.53$, $df = 48$, $p < .05$) but indistinguishable from stories from patients with no PDs (two-tailed t-test, $t = -1.32$, $df = 34$, $p = .20$).

Although stories from patients with a simple PD diagnosis appeared to be longer than stories from those with a complex PD diagnosis, this difference was not significant (two tailed t-test, $t = 1.80$, $df = 24$, $p = .09$).

In summary, the word lengths of stories from different groups were as follows:

Table 12.1: Summary of transcript length by group

	Transcripts from:		
	Clinicians	Patients: No PD	Patients: at least one PD
Number of transcripts	23	13	26
Mean word length (<i>sd</i>)	1340 (540)	1610 (650)	990 (440)

It was possible that another factor was influencing both story length and PD diagnosis; the likeliest factor for which data were available was felt to be level of education achieved. On inspection it did appear as though there was an interaction. Stories from patients with at least one PD diagnosis seemed distinctly longer in those patients who had continued in education after the age of 16. Patients with no PD diagnosis had no such difference. Figure 12.3 below illustrates these findings.

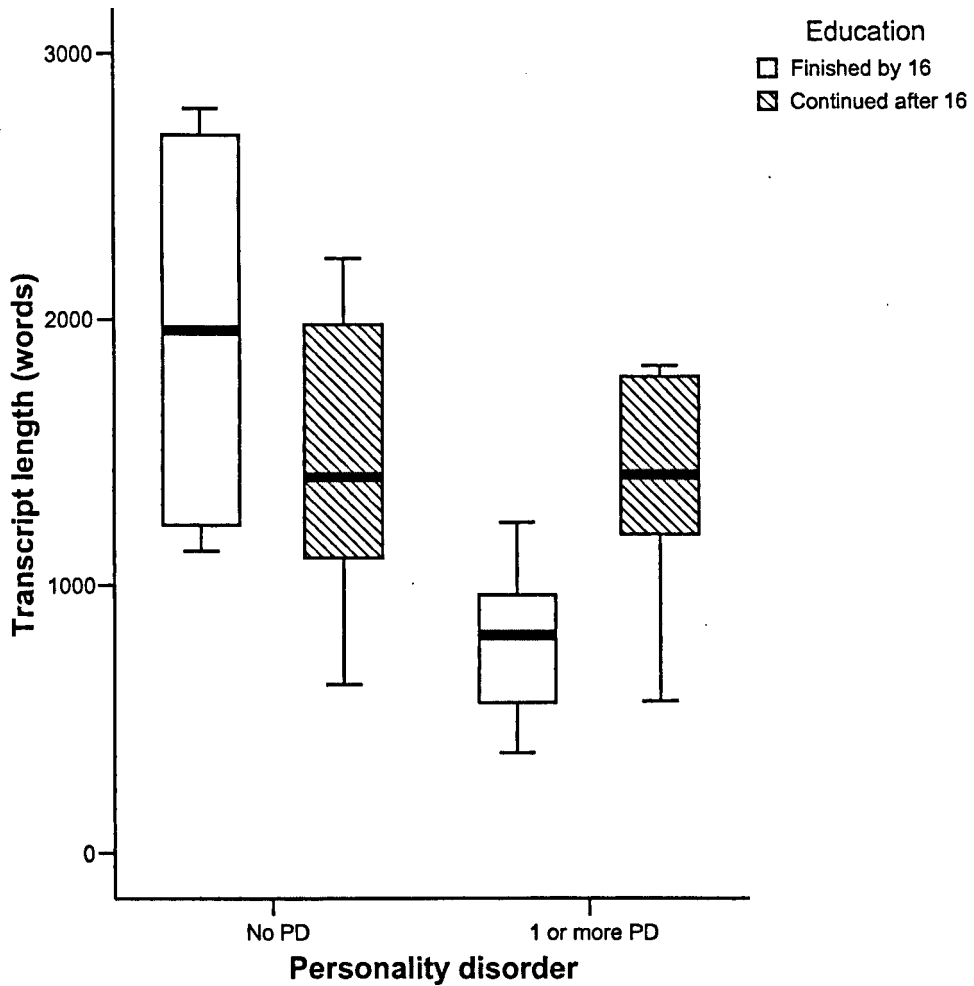


Figure 12.3: Boxplot of story length by level of education and personality disorder diagnosis.

The raw correlation between complexity of PD diagnosis and length of story transcript appeared to be significant (two-tailed test, $r = -.54$, $n = 38$, $p < .01$).

When education level was controlled for the correlation declined but remained significant ($r = -.41$, $df = 35$, $p < .05$).

A multiple regression calculation was carried out to see which of education level and complexity of PD diagnosis was the better predictor of story length.

Education accounted for 30% of the variance in story length and complexity of PD diagnosis a further 12%. This result suggests that the length of story is more

influenced by the level of education reached than the degree of personality disorder.

It was thought possible that IQ might be a possible confounding factor here, so a literature search was undertaken to see whether story length and IQ were associated in any of the well-researched projective approaches such as the Rorschach and the TAT. One TAT reference (Holt, 1992, originally published 1958) referred in general terms to story length as an interesting variable, but gave no indication of what it might be associated with. No other references were found and approaches to some experts in the field yielded no more data.

One other feature of story transcripts was investigated, following on from an observation in previous clinical practice. It had been noted that in 6PSM sessions with patients, although their initial stories tended often to be short, subsequent questioning produced a large amount of further detail. In contrast, 6PSM sessions done with clinicians (eg students, attendees at training courses) tended to have much more detailed initial stories with the result that the subsequent questioning was shorter. Each story was divided into two, such that:

Total story length = Length of first telling + Length of responses to questions

A new variable designated the *Response/story ratio* was then calculated thus:

Response/story ratio = Length of responses to questions + Length of first telling

The differences in the Response/story ratios of stories from different groups are given in the Table 12.2:

Table 12.2: Response/story ratios of stories from different groups

Stories from:	Response/story ratio				
	<i>n</i>	Mean	<i>SD</i>	Min	Max
Clinicians	24	2.2	0.7	0.8	3.6
Patients without BPD diagnosis	21	3.8	2.1	1.8	8.2
with BPD diagnosis	19	3.5	1.7	1.3	8.3

Clinicians, on average, seem to use a little over twice as many words in their responses to questions as they did in their original stories. Patients on average use between three and four times as many words, and for some the ratio is over eight. In the most extreme case, a patient story of 59 words was followed by 492 words in response to questions. There was no significant difference in Response/story ratio between patients with and without a BPD diagnosis ($t = .49, df = 38, p > .05$). Clinicians and patients however differed significantly in their Response/story ratios ($t = -4.46, df = 56.44, p < .001$, equal variances not assumed). This is illustrated overleaf (Figure 12.4).

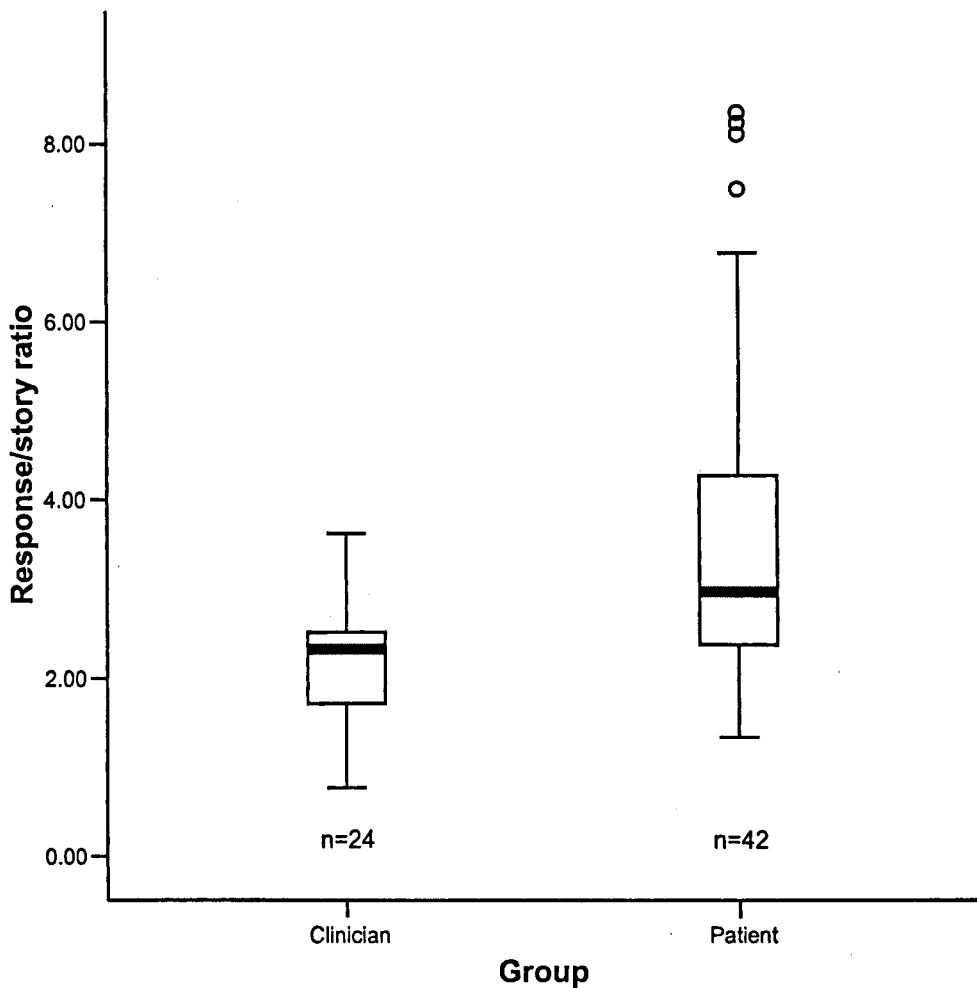


Figure 12.4: Boxplot of Response/story ratio of stories from clinicians and patients. Circles represent outlier stories with a response/story ratio more than three times the semi-interquartile range above the median value.

Taking all the word length data together, the results suggest that clinicians and patients without a BPD diagnosis produce stories of similar overall length, significantly longer than those produced by patients with a BPD diagnosis. In the ratio of initial story to question responses however, both patient groups are similar to one another in producing proportionately shorter stories and longer question responses than clinicians.

12.3 COMPUTER BASED TEXT ANALYSIS

All the computer based text analysis undertaken followed the same basic premise. Each program is simply a counting routine that identifies the number of times any given word is used in the story. Pre-existing dictionaries containing categories of words were then employed, with the number of words in each category being calculated for each story. The dictionaries were all ones which have been used in the past for the investigation of the psychological state of the writer of a piece of text (or the speaker, in the case of transcribed text.)

The programs used, with the dictionaries employed, were as follows:

Table 12.3: Text analysis programs and dictionaries used

Program	Dictionaries	Reference
General Inquirer	Harvard-IV, Lasswell	(Stone, Dunphy, Smith, & Ogilvie, 1966)
WordStat	Regressive Imagery Dictionary (RID)	(Martindale, 1975)
Not known (analysis conducted elsewhere)	Linguistic Inquiry and Word Count (LIWC)	(Pennebaker & King, 1999)

In addition, WordStat was used to construct a custom built dictionary for this study, as described below.

12.3.1 GENERAL INQUIRER

The General Inquirer (Stone *et al.*, 1966) was one of the earliest text analysis programs to be developed. The categorisation system of the Harvard-IV and Lasswell dictionaries has 183 categories into which words are counted, some of which overlap considerably. The General Inquirer (GI) computer program

produces output in the form of the number of words that fall into each category per 100 words of the text. Large categories of commonly occurring words may reach counts approaching 20 words per 100; in other words nearly a fifth of the text will be words of that category. Smaller categories, or those with words that occur more rarely, will have smaller counts. With this body of 67 transcribed stories the 183 categories occurred with the following mean frequencies:

Table 12.4: Frequency of occurrence of 183 General Inquirer categories

Mean word frequency	<1%	1-2.9%	3-4.9%	5-6.9%	7-8.9%	9%+
Number of categories	137	20	7	4	3	2

The majority of GI categories occurred at a frequency of less than 1 word per 100; only 36 exceeded this. These 36 were further investigated to see whether they had any relationship to the clinical data relating to the stories' authors. The 137 categories that appeared less frequently were not investigated further, because of the danger of spurious significances being detected when such a large number of possible variables was being analysed.

12.3.1.1 Constructing a measure of borderline personality disturbance from GI data

The aim was to identify a subgroup of the GI categories that were characteristic of patients with a diagnosis of Borderline Personality Disorder. It would have been possible simply to test the correlation of each GI category with the number of BPD diagnostic criteria met during the SCID-II interview. The categories that correlated significantly could be assembled into a scale that could then be tested to see how well it distinguished those with a BPD diagnosis from others.

However this would have used the SCID-II data twice; once to construct a scale and again to test it. This would have been a circular process with little chance of failure and consequently not very impressive if successful.

Instead one of the other concurrent instruments was used to initially select GI items. The CORE Risk subscale was felt likely to be a good predictor of the level of Borderline Personality disturbance. The correlation was calculated between each of the 36 GI categories and the CORE Risk subscale of the story's author. The distribution of the CORE Risk z-scores was markedly different from the normal distribution (see Figure 6.5 below) and so the Spearman rank correlation coefficient was used. A two-tailed test was used because no predictions were being made about the direction of any correlation.

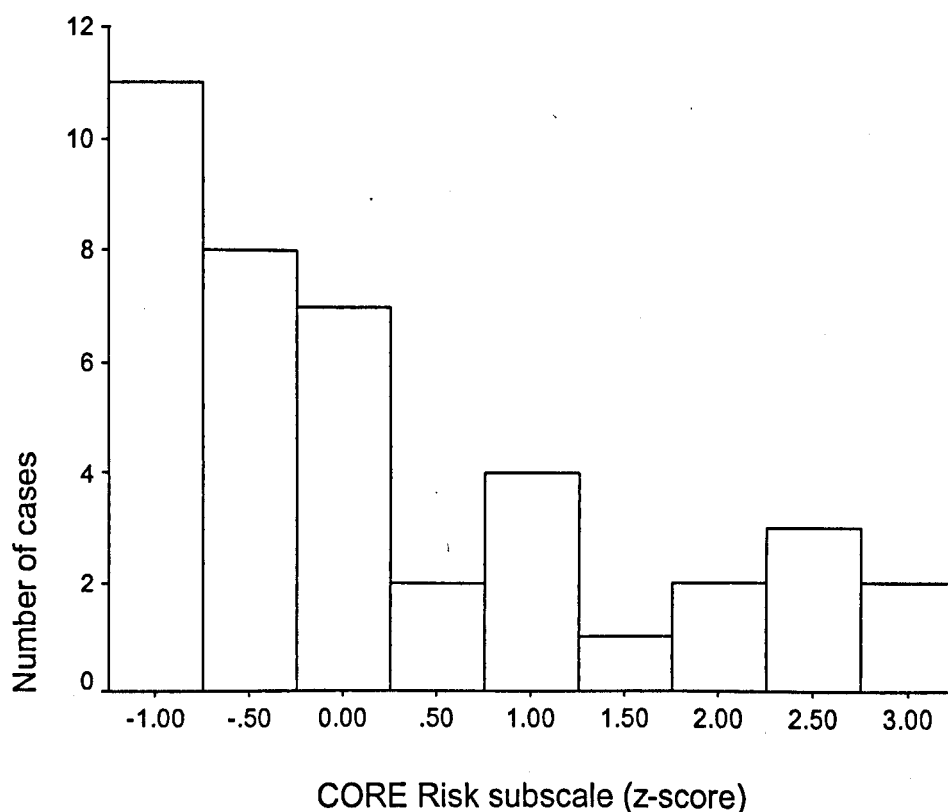


Figure 12.5: CORE Risk subscale z-scores of story authors, demonstrating departure from the normal distribution.

Those GI categories that correlated significantly with the CORE Risk subscale score of the story's author were assembled into a scale by simple addition and the resulting scale tested to see how well it predicted a diagnosis of BPD as reported by the SCID-II. The GI categories involved were:

Table 12.5: General Inquirer categories correlating significantly with CORE Risk subscale

Category Label	Category Description	Examples of words	<i>r</i>
SPACE	Words indicating a consciousness of location in space and spatial relationships	Acre, beyond, enclose, high, middle, position, thin, yonder.	.36*
TIMESPC (Time-space)	Words relating to space and time	Afternoon, central, elapse, height, instant, outer, past, route, tall.	.63***
OVRST (Overstatement)	Words indicating emphasis in realms of speed, frequency, quantity accuracy, size, intensity, certainty and extremity	Abominable, delirium, great, justify, matchless, perpetual.	-.35*
CAUSAL	Words denoting presumption that occurrence of one phenomenon is necessarily preceded, accompanied or followed by the occurrence of another.	Accident, because, effect, likely, reason, result, why	.35*
HU (Human)	Words referring to humans, their social roles and categories.	Alien, companion, family, grandchild, Mrs, prisoner, role, woman, youngster	-.41**

Note. In the case of all the correlations reported above, $n = 40$.
* $p < .05$, two-tailed, ** $p < .01$, two-tailed, *** $p < .001$, two-tailed.

Of these five categories, those where $p < .01$ were retained for assembly into a scale. This significance level was chosen because there were 36 categories being correlated with the CORE Risk subscale. A significance level of $p < .05$ would have been likely to produce at least one Type I error.

The two categories were assembled into a scale in the following way. For each story, the number of words per hundred in the positively correlated TIMESPC category was noted and the number of words in the negatively correlated HU category was subtracted from this. The resulting score was called the GI-B score (General Inquirer - Borderline Personality Disorder) and recorded for each story. An independent samples *t*-test was then carried out to see how stories from patients with a diagnosis of BPD differed from other stories; GI-B scores from the two groups were significantly different ($t = -3.25, df = 38, p < .01$).

Some patients completed two stories, approximately one month apart; it was therefore possible to compare the GI-B scores obtained on the two occasions and these were found to be moderately correlated ($r = .44, n = 18, p < .05$). This is demonstrated in the scatter plot overleaf (Figure 12.6). In this figure each data point represents a patient, with the two axes showing the GI-B scores of the two stories they produced.

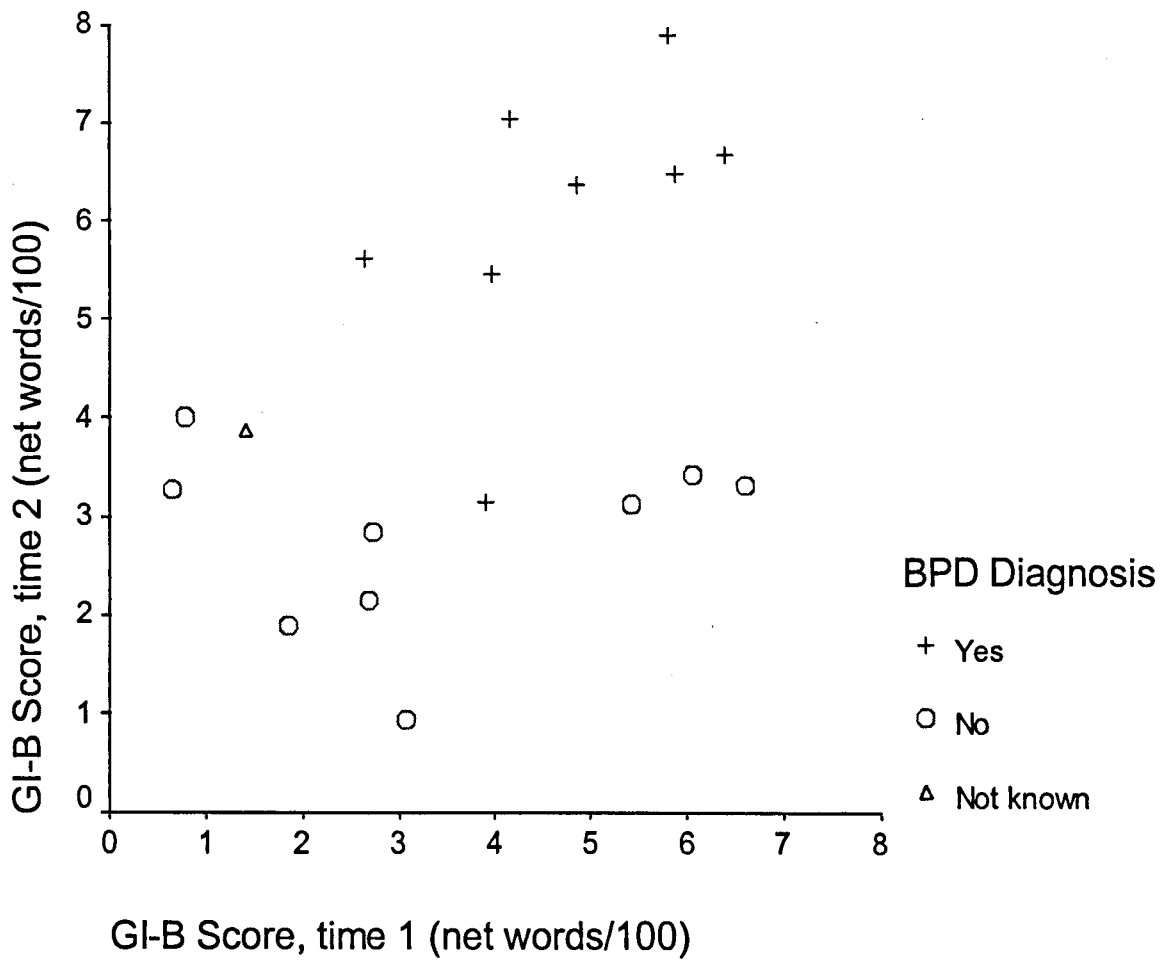


Figure 12.6: Stability of GI-B score over time

A Receiver Operating Characteristics (ROC) curve was drawn to test the sensitivity and specificity of this newly derived GI-BPD score. The 40 stories from patient participants where SCID-II diagnostic data were available were used. The ROC curve was interpreted according to the guidelines in Streiner and Norman (1995: p96-101). The optimum cut-off point is that part of the ROC curve closest to the top left hand corner of the graph, thus maximising both sensitivity and specificity.

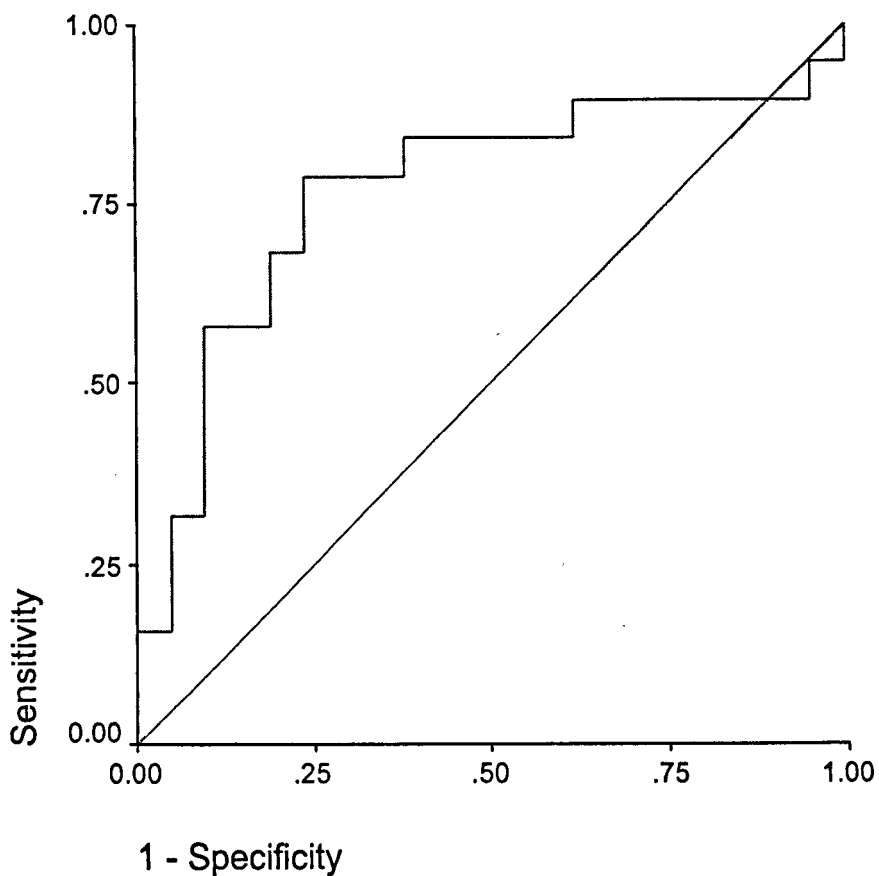


Figure 12.7: ROC curve for derived GI-B score

At the optimum cutoff point (a GI-B score of 3.67 or more) the scale had a sensitivity of 79% and a specificity of 76%. This is amplified in the contingency table below, suggesting that the GI-B score discriminates reasonably well between stories from patients with a BPD diagnosis and others.

Table 12.6: Predicted and actual authorship of stories, as estimated from GI-B score

Actual status from SCID-II interview	Predicted status from GI-B Score	
	Author with BPD diagnosis	Author with no BPD diagnosis
Author with BPD diagnosis	17	4
Author with no BPD diagnosis	5	14

12.3.1.2 Comparison of GI-B score with concurrent measures

A comparison with other measures was made by examining the correlation between the GI-BPD score from patient stories and the CORE subscales, the IIP mean, the number of SCID-II criteria met and the total number of PDs diagnosed. The correlation with the CORE Function subscale was not significant ($r = .24, n = 40, p = .14$), but the other seven measures were significantly correlated. As expected the correlation with the CORE Risk subscale was the most significant ($r = .55, n = 40, p < .001$).

It was to be expected that the correlation with the CORE Risk subscale would be high, because the GI-B scale had been constructed from correlations with the CORE-R variable. However the scatter plot of GI-B scores against CORE Risk z-scores was still striking (Figure 12.8 overleaf).

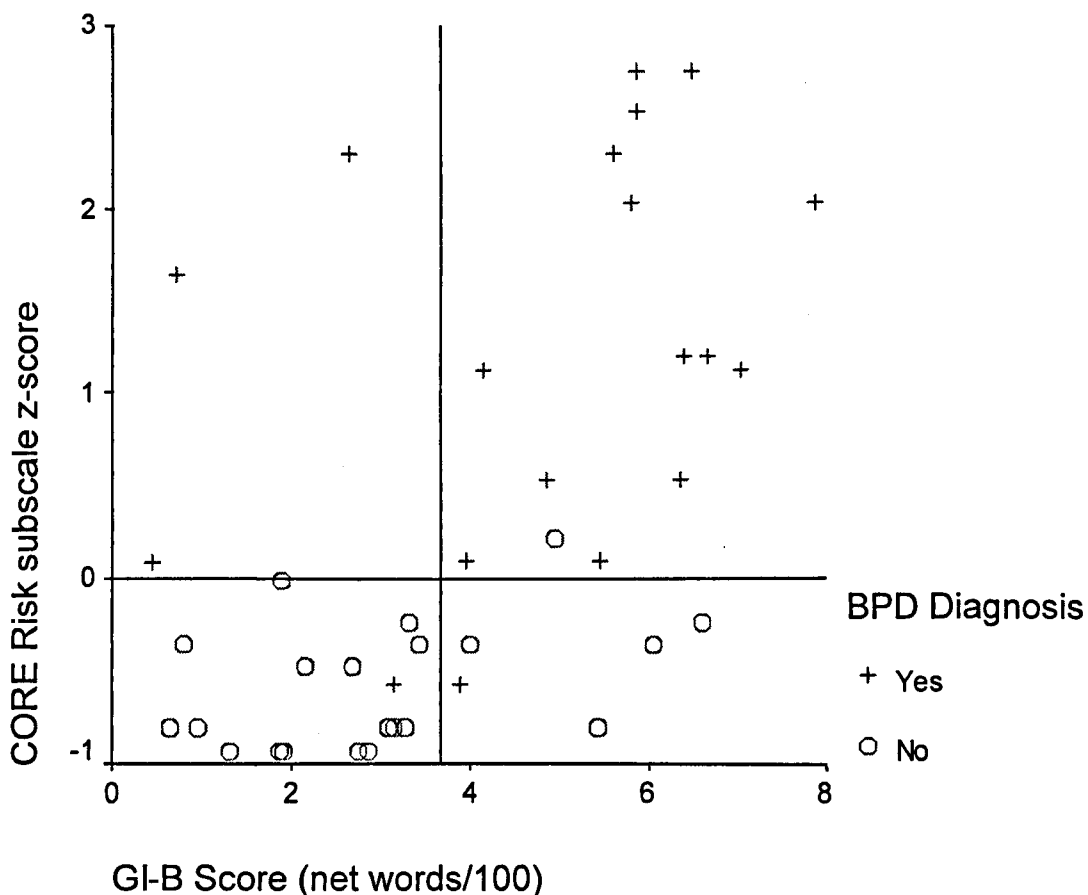


Figure 12.8: Scatter plot of story GI-B score against CORE Risk z-score of author

A line at the GI-B cut-off point (GI-B = 3.67) has been drawn in the above figure, along with a line representing the normative mean for the CORE Risk subscale (z = 0). The contrast between the lower left quadrant (dominated by stories from patients without a BPD diagnosis) and the upper right quadrant (dominated by those with such a diagnosis) is clear to see. The other two quadrants remain relatively empty, suggesting a strong link between the derived GI-B score, the CORE Risk subscale and a diagnosis of Borderline Personality Disorder. Thus data from computer-based text analysis, patient self-report and structured clinical interview appear to triangulate well.

12.3.1.3 Comparison of GI data from stories only with full transcript

As already explained, the transcripts analysed above were the complete transcripts including the story as first told uninterrupted and the teller's answers to the standard amplifying questions. The GI results from these full transcripts were compared with the GI categories applied to the story alone, without the amplifying answers. The length of text analysed was much shorter, but it was thought possible there might be differences between the kind of language used spontaneously in the first telling and the language used in response to the scripted questions. Possibly the former was purer and less influenced by the listener's questions (although the same questions were asked of each teller).

The same process outlined above was undertaken for the GI categories as applied to the shorter first telling. Of the 46 GI categories, only two proved to be significantly correlated with the CORE Risk subscale, one at the level $p < .01$ and one at the level $p < .05$. Remembering that eight categories had been correlated at the level $p < .05$ or better when analysing the full transcript, it was felt that these two correlations out of 46 analyses undertaken were much more likely to be spurious and to have arisen by chance.

The shorter transcript of the story alone was not therefore further analysed, assuming that the larger word count of the full transcripts would make it more likely that significant differences would be detected, and less likely that these would arise by chance.

12.3.2 LINGUISTIC INQUIRY AND WORD COUNT (LIWC)

LIWC is a categorisation system that has been developed explicitly to try and reveal information about the psychological state of the writer of a piece of text (Pennebaker & King, 1999). The same process was used with the LIWC categories as with the GI categories to try and identify any LIWC categories that distinguished between stories from patients with a BPD diagnosis and those without.

There are 72 categories in the LIWC dictionary, but only those with a mean frequency greater than 1 per 100 words across all stories were investigated. This was a group of 30 LIWC categories. The frequency of occurrence of words in these categories was correlated against the author's CORE Risk subscale score. A lower significance threshold of $p < .02$ was chosen rather than $p < .05$ because with 30 correlations being performed there was a risk of a Type I error. The following categories were significantly correlated:

Table 12.7: Linguistic Inquiry and Word Count categories correlating significantly with CORE Risk subscale

Category label	Description of category	Examples of words	<i>r</i>
SPACE	Words to do with relative positions in space	Around, over, up	.56***
CERTAIN	Words to do with the cognitive process of certainty	Always, never	-.43**
MOTION	Words to do with the motion through space	Walk, move, go	.39†

Note. In the case of all the correlations reported above, $n = 40$.
 † $p < 0.02$, ** $p < 0.01$, *** $p < 0.001$, two-tailed.

Once again, a single LIWC Borderline score, labelled LIWC-B, was produced by summing the word counts in the two positively correlated categories and subtracting the word count in the negatively correlated category. The LIWC-B score proved to have a significant test-retest reliability ($r = .62, n = 18, p < .01$).

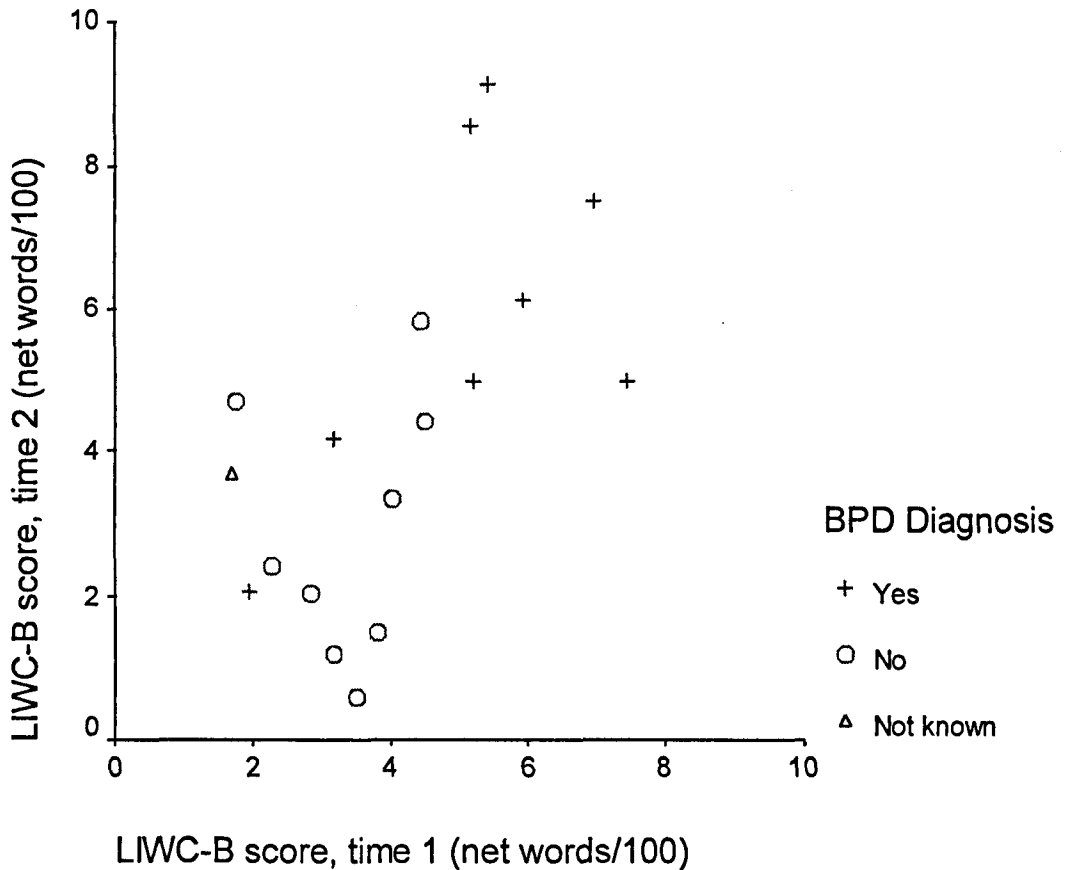


Figure 12.9: Stability of LIWC-B score over time

The discriminant ability of the LIWC-B was further investigated by comparing the mean LIWC-B score of stories produced by patients with a BPD diagnosis and patients without. The former stories scored significantly higher ($t = -3.52, df = 38, p < .01$). A ROC curve was drawn to identify the optimum cutoff point for discriminating between the two sorts of stories, shown below.

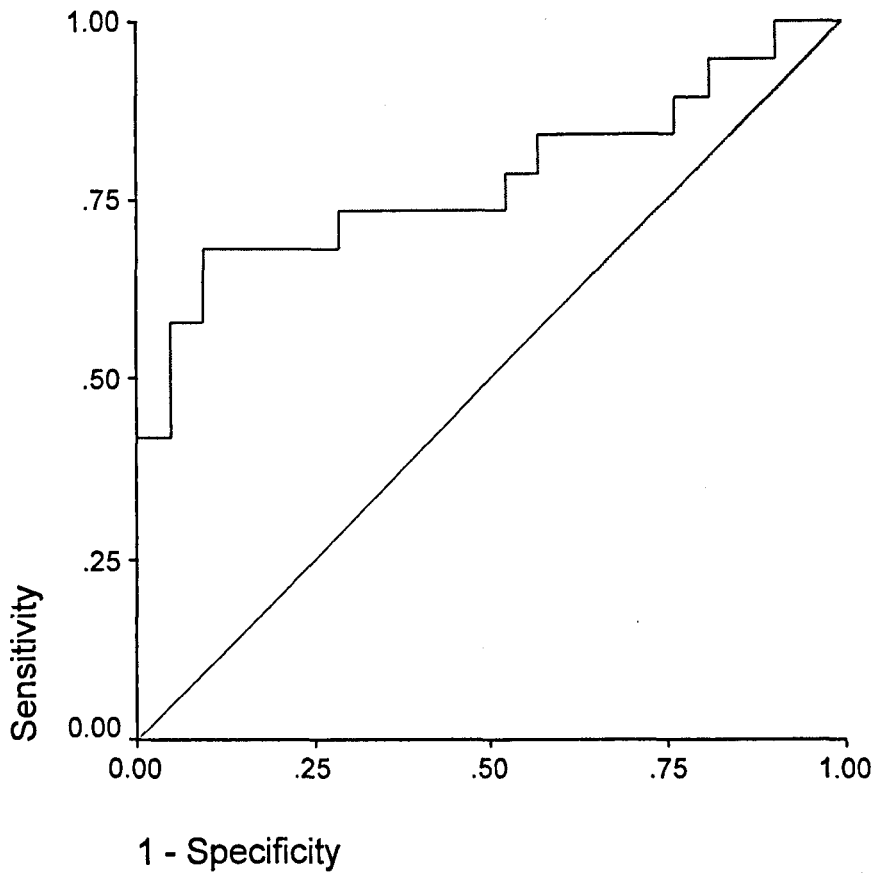


Figure 12.10: ROC curve for derived LIWC-B score

This suggested a cutoff point of 4.84 for the LIWC-B score which results in a sensitivity of 68% and a specificity of 91%, as amplified in the table below.

Table 12.8: Predicted and actual authorship of stories, as estimated from LIWC-B score

Actual status from SCID-II interview	Predicted status from LIWC-B Score	
	Author with BPD diagnosis	Author with no BPD diagnosis
Author with BPD diagnosis	15	6
Author with no BPD diagnosis	2	17

12.4 THE REGRESSIVE IMAGERY DICTIONARY

Another categorisation system is the Regressive Imagery Dictionary (Martindale, 1975). The RID is a categorisation system that seeks to identify primary process and secondary process imagery. The former is seen as oriented inwards towards the self and subjective experience and the latter outwards towards the outside world, events and objective reality (Stigler & Pokorny, 2001: p416). The RID also counts categories such as emotions (glory, anxiety) and other categories (novelty, male/female roles etc).

The 38 stories from patients were compared to see which RID categories differed significantly between those with and those without a BPD diagnosis. Three RID categories seemed to differ significantly; patients with a BPD diagnosis used more words from two categories, and fewer from a third. The three categories were:

Table 12.9: Regressive Imagery Dictionary categories correlating significantly with CORE Risk subscale

Rid category/ Subcategory	Description	Examples of words	<i>r</i>
Primary process/ Concreteness	Words indicating concrete physical description	Across, behind, distant, flat, near, side, west, wide	*.36
Others/Negation	Words indicating negation	Absence, blank, never, nobody, not, nowhere	*.36
Secondary process/Abstract thought	Words indicating abstract thought processes	Analyse, believe, calculate, decide, guess, real, suppose, weigh	*-.33

Note. In the case of all the correlations reported above, $n = 40$.

* $p < 0.05$, two-tailed.

Once again a RID Borderline vocabulary score (RID-B) was constructed by adding the word counts in the Concreteness and Negation categories and subtracting the word count in the Abstract thought category. In this case, because the RID categories are mutually exclusive, it was known that there would be no double counting.

The RID-B scores at times 1 and 2 correlated moderately ($r = .43, p < .05, n = 18$, one-tailed), and the RID-B score was significantly different for stories from patients with and without a diagnosis of BPD ($t = -2.41, df = 38, p < .05$).

However the absolute value of t and the significance level were lower than those for the GI-B and LIWC-B scores. The test-retest stability is demonstrated in Figure 12.12.

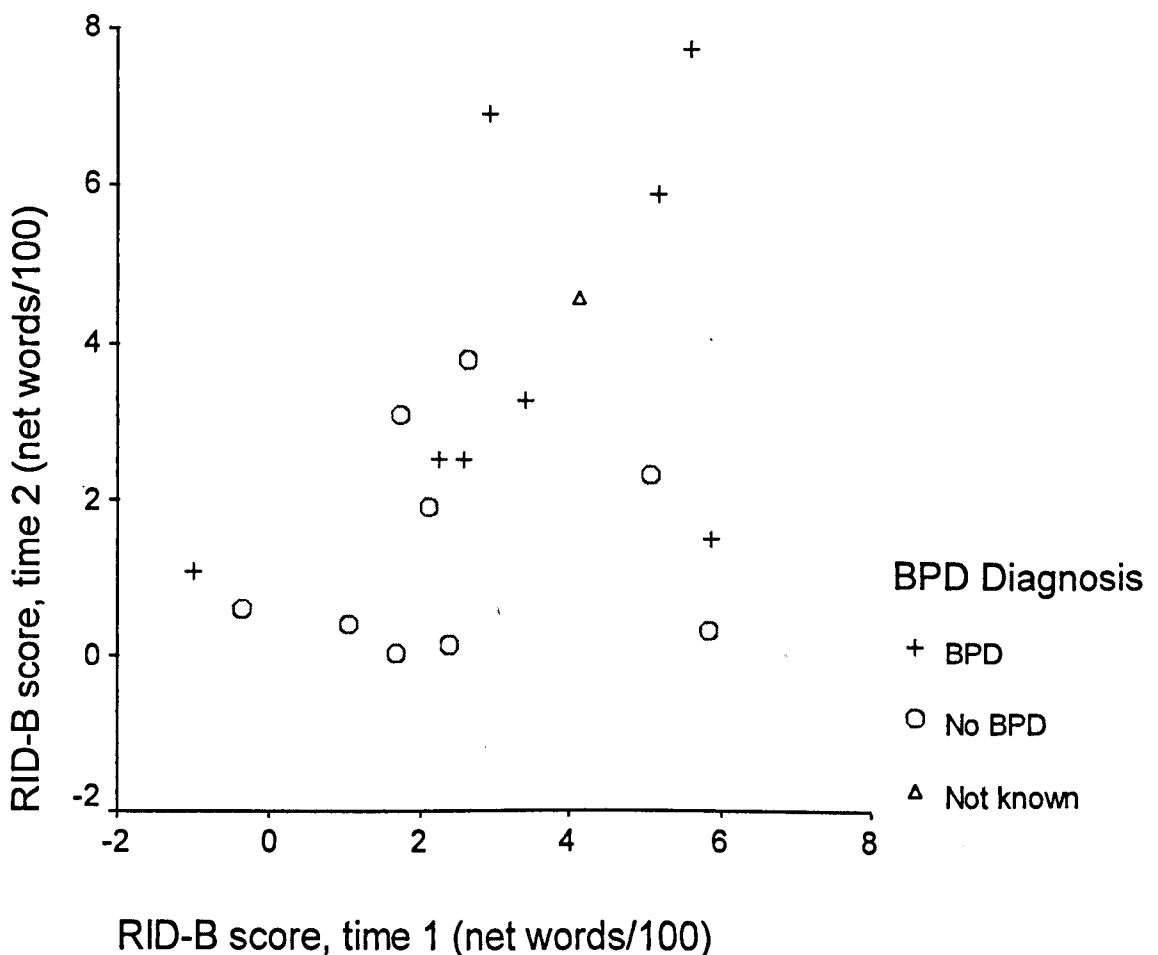


Figure 12.12: Stability of RID-B score over time

A ROC curve shown in Figure 12.13 provided an optimum cut-off point of 2.45, giving a sensitivity of 74% and a specificity of 71%.

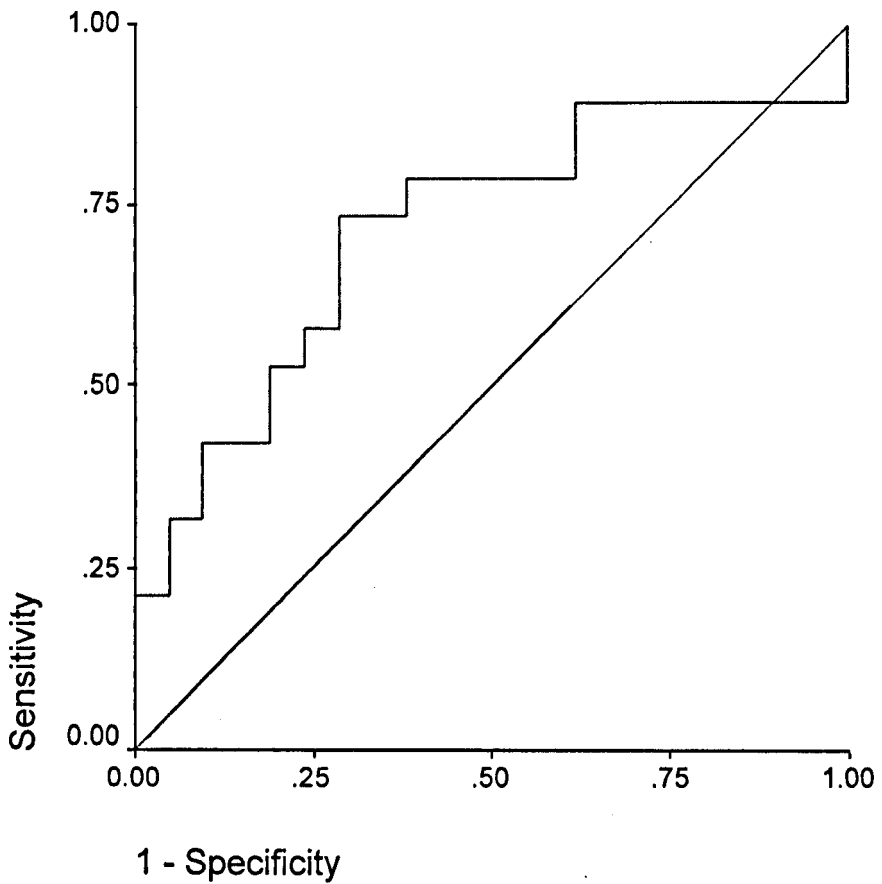


Figure 12.13: ROC curve for derived RID-B score

This cut-off resulted in the following contingency table for the identification of the authorship of stories (Table 12.10 overleaf):

Table 12.10: Predicted and actual authorship of stories, as estimated from RID-B score

Actual status from SCID-II interview	Predicted status from RID-B Score	
	Author with BPD diagnosis	Author with no BPD diagnosis
Author with BPD diagnosis	16	5
Author with no BPD diagnosis	6	13

Finally, the correlations with the concurrent measures were calculated, as had been done for the GI-B and LIWC-B scores. The RID-B score correlated significantly ($p < .01$) with all the CORE scales and subscales, the IIP mean, the number of SCID-II criteria met and the total number of Axis II disorders diagnosed.

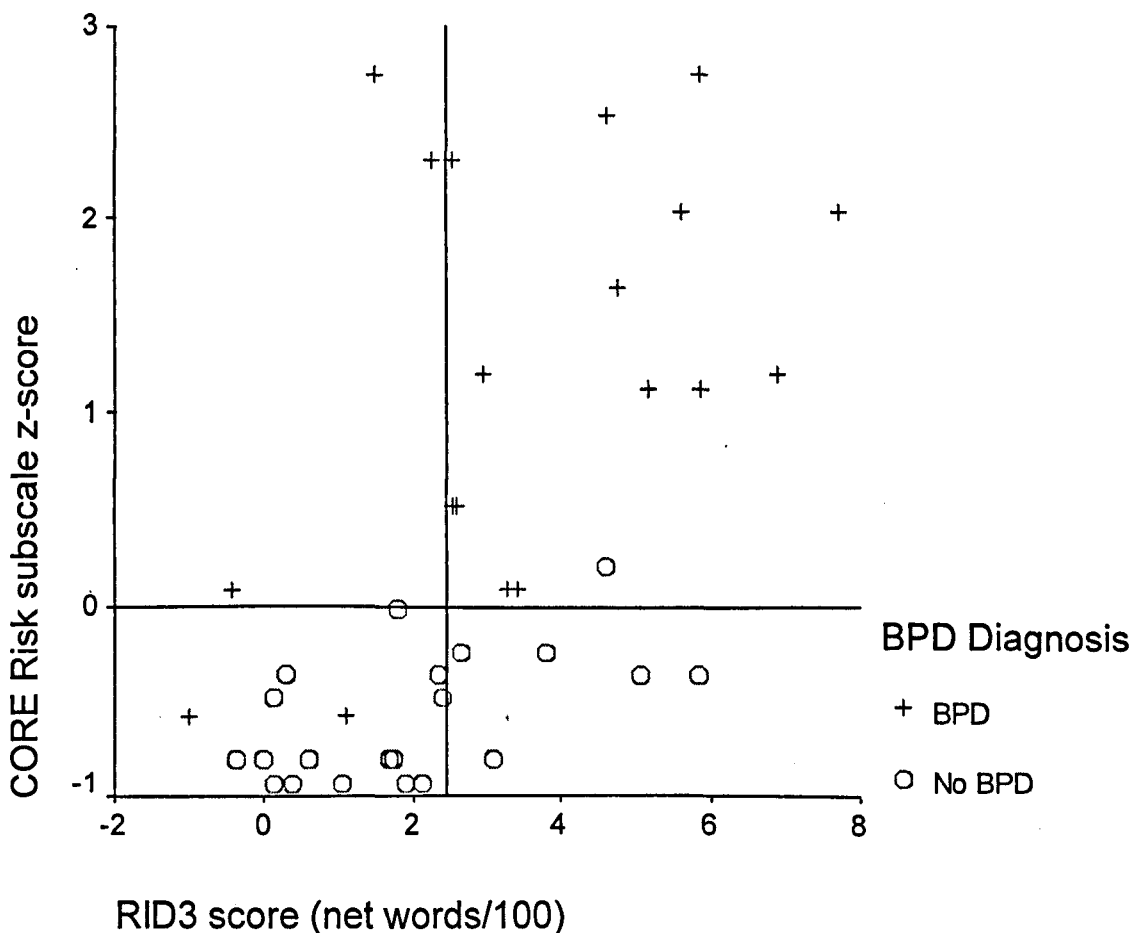


Figure 12.14: Scatter plot of story RID-B score against CORE Risk z-score of author

12.5 CONSTRUCTION OF NEW CATEGORIES

The text analysis program used to calculate the RID data is known as WordStat (Péladeau, 1998). WordStat can also be used to construct categories using a dictionary-building tool, following principles of dictionary-building (Bengston & Xu, 1995). One set of word categories that has been suggested was relevant to 6PSM is the BASICPh system (Lahad & Ayalon, 1993). Lahad and Ayalon were using 6PSM to assess coping styles in school pupils in northern Israel, and they inspected stories for evidence of coping in six categories:

- B= Beliefs: Belief in one's own capacity, faith in others, knowledge of good and evil, moral and ethical values, existential issues.
- A= Affect: Coping through expression of emotion, catharsis, feelings.
- S= Social: Using social networks or relationships, gaining support from others, asking for or receiving help.
- I= Imagination: Coping by imaginative means, using magical, novel or unexpected strategies.
- C= Cognitive: Problem-solving, logical, rational, real-world methods of coping.
- Ph= Physical: Coping by physical means; fighting, running away, building, making.

No pre-existing dictionary exists of these six categories, although examples of the kinds of words and concepts used have been given (Lahad & Ayalon, 1993). These examples were used to start the iterative process of building a dictionary. The process starts by entering some initial words into WordStat and

categorising them into one of the six BASICPh categories. This was done by taking words described by Lahad and Ayalon in their article. WordStat then expands the dictionary by offering synonyms of the given words, as well as other words derived from inflections of the roots provided. For example, if 'think' and 'solve' are the two root words in the Cognition category, WordStat suggests the following other words to go into the same category:

Table: 12.11: Synonyms and inflected forms generated from the example roots *think* and *solve*.

Synonyms	Inflected forms	
Answer	Solvency	Thinks
Decipher	Solvencies	Thinking
Decode	Solvent	Thinker
Determine	Solvently	Thinkers
Explain	Solvents	
Figure out	Solved	
Interpret	Solves	
Resolve	Solver	
Unravel	Solvers	
Untangle	Thinkable	

Words may be deleted from the lists if they are felt to be irrelevant (for example, the words from *solvency* to *solvents* might be excluded) and the dictionary building process can be gone through again, using the newly added words as roots. This quickly allows a category to be developed from a small nucleus of root words. In addition, the WordStat program allows the words from the corpus under scrutiny to be counted, and all those words which have not yet been categorised can be displayed in order of decreasing frequency. Thus a word occurring frequently in the stories, but not yet classified in the 6 categories, can be identified. The most frequent words in the 67 stories in this study were *the*, *and*, *he* and *to*. Between them these four words occurred 11,902 times in the 80,685 words of the stories. Beyond these commonplace words however it was

possible to identify other words fitting the BASICPh categories; for example *people* (253 occurrences) was placed in the Social category and *know* (193 occurrences) into Cognition.

The potential dictionaries thus assembled were likely to contain words that were not used, in practice, in the predicted ways. To check for this a Key Word In Context (KWIC) analysis was carried out. WordStat permits KWIC analysis by listing every instance of each word in a given category in context. An example is given below. The key word is shown in bold and its context (the rest of the sentence in which it appears) is given. The example word, *race*, appeared 4 times in the 67 stories.

*I don't think there's anything different, the only difference is there's a church there and the family are a different **RACE**.*

*The whole **RACE** are bored.*

*And the TNT, that's just some wastage that someone out of the human **RACE** has gone and dumped in the middle of nowhere to get rid of.*

*It's a nice place, it's full of greenery, it's been untouched, it hasn't been spoiled by the human **RACE** or anything and it's a very nice place to be, nice and warm and he's got everything he needs.*

The word *race* had been ascribed to the Physical category because of the possible connotations of physical competition, speed and exertion. However on inspection all the instances of *race* had to do with connotation of human (or other) race. The word was therefore reassigned to the Social category. If a word could not be assigned to a single category it was deleted, rather than being assigned to more than one category.

In addition, for a word to be assigned to a category it had to accurately capture the sense of that category a minimum of 80% of the time, otherwise it was

removed. This is a commonly used rule of thumb in text analysis (Bengston & Xu, 1995: p10). For example, the words *felt* and *feel* were both initially assigned to the Affect category. However a KWIC inspection revealed that while *feel* was almost always used in the context of emotions, the same was not true of *felt*. Nearly half of the instances of *felt* had to do with holding an opinion or belief rather than experiencing an emotion. *Feel* was therefore retained in the Affect category, and *felt* removed.

Once the six categories and their component words had been constructed and validated the values of the six categories were compared with the CORE Risk subscale by testing correlation with the CORE Risk subscale, as explained above for the General Inquirer and other data. No significant association was found between any of the six BASICPh categories and the CORE Risk subscale.

The other items of the concurrent measures were also compared against the six BASICPh categories, to see whether there were any other associations.

In the CORE scale and subscales, only the Social category from BASICPh had any significant correlation/ It was positively correlated with the CORE Wellbeing, Problems and Function scales as well as the CORE non-risk items total which summarises these. The strongest correlation was with the CORE non-risk items total ($r = .35, n = 38, p < .05$). Among elements of the other concurrent measures, only the Physical category from BASICPh showed any correlation, with the IIP mean score ($r = .32, n = 38, p < .05$).

The scores in the six BASICPh categories were also tested for test-retest reliability. The results were:

Table 12.12: Test-retest stability of BASICPh categories

BASICPh category	<i>r</i> (two-tailed)
Belief	*.50
Affect	** .59
Social	.37
Imaginative	.27
Cognitive	*.47
Physical	** .57

* $p < .05$, ** $p < .01$

None of the six BASICPh categories appeared to distinguish well between stories from patients with and without a BPD diagnosis. In no case were the mean category scores significantly different, the closest to significance being that for the BASICPh Physical category ($t = -1.36$, $df = 24.54$, $p = .19$).

12.6 SUMMARY OF RESULTS OF COMPUTER BASED TEXTUAL ANALYSIS

The newly constructed BASICPh dictionary and its six categories did not yield any useful data to identify stories from patients with a BPD diagnosis. The three existing dictionaries, the General Inquirer (GI), Linguistic Inquiry and Word Count (LIWC) and the Regressive Imagery Dictionary (RID) did however yield scores that could usefully make such distinctions.

Table: 12.13: Summary of performance of scores derived from General Inquirer, Linguistic Inquiry and Word Count, and Regressive Imagery Dictionary

Score	Test-retest reliability	Discriminant ability	
		Comparison of means of stories (BPD vs. no BPD)	Sensitivity / Specificity at optimum cut-off
GI-B	$n = 18, r = .44^*$	$t = -3.25^{**}, df = 38$	79% / 76%
LIWC-B	$n = 18, r = .62^{**}$	$t = -3.52^{**}, df = 38$	68% / 91%
RID-B	$n = 18, r = .43^*$	$t = -2.41^*, df = 38$	74% / 71%

* $p < .05$, ** $p < .01$

Of the three measures created, that derived from the LIWC had the best test-retest reliability, the highest absolute t -value for the comparison of means and the best specificity at optimum cut-off. The GI-B score had the best sensitivity at optimum cut-off.

12.7 QUANTITATIVE TEXT ANALYSIS: OTHER FEATURES

There were some other notable features of the stories produced, beyond those relating to personality disorder. One feature that had been noted before was the preference for male main characters to be chosen, both by male and female tellers. This pattern was repeated among the patients in this study. However, a significant proportion of women - although still a minority - did choose female main characters.

Table 12.14: Gender of main character and teller

Gender of author	Gender of main character		Total
	Same as teller	Different from teller	
Male	18	1	19
Female	^a 8	40	48
Total	26	41	67

^aOne participant was a male-to-female transsexual, classified here as female according to her preference. Both her stories had female main characters.

A chi-squared test on this contingency table revealed a significant difference ($\chi^2 = 34.94$, $df = 1$, $p < .001$). Post-hoc partitioning of the table suggested that the greatest contribution to the value of chi-squared came because no male tellers chose female main characters (one male teller chose a main character of unspecified gender.) This result suggests that a story with a female main character is highly likely to have been written by a woman. However, a story with a male main character cannot be said to be predictive of a male author.

As a consequence of the preference for a male main character, there was a great preponderance of language with a masculine reference (him, he, his) and this was detected in the text analysis. The LIWC has two categories, one each for words with a male and female connotation. Of the 19 stories from male subjects, all had more words in the former category than the latter. Perhaps more surprisingly, of the 48 stories from female subjects 40 of them also had more words in the male category than the female category. Only eight of the 48 stories produced by women had more words in the female category than the male category. Interestingly, one of these was written by a pre-operation male to female transsexual, who was classified as female for this study because this was her preferred description of herself.

The types of main character selected were also investigated. Once all 67 stories were transcribed, the nature of the main characters was inspected and a simple coding scheme devised to describe them. Main characters were divided into five types: human, fantasy human, animal, fantasy creature and inanimate. A further breakdown of these is given overleaf:

Table: 12.15: Classification of character type and subtype chosen

Character type and Subtype	<i>n</i>
Human	17
Man	13
Girl	2
Boy	1
Woman	1
Fantasy Humanoid	11
Alien	2
Blobby man	2
Humanoid monster	2
Angel	1
Clanger	1
Elf	1
Ghost	1
Hobbit	1
Animal	24
Cat	7
Dog	4
Bird	2
Giraffe	2
Pig	2
Tortoise	2
Armadillo	1
Camel	1
Crocodile	1
Horse	1
Squirrel	1
Fantasy Animal	5
Dragon	2
Unicorn	2
Indeterminate creature	1
Inanimate	10
Star	2
Teddy bear	2
Car	1
Computer	1
Cottage	1
Doll	1
Mobile phone	1
Orange (fruit)	1

The five main categories were investigated to see whether there were any differences in the choices made by different subgroups. The most striking difference was in the relative frequency of choice of a human main character between clinicians, those patients with a diagnosis of BPD and those without.

The contingency table below demonstrated significant differences between the groups ($\chi^2 = 10.89$, $df = 2$, $p < .01$).

Table: 12.16: Character choice of authors by group and diagnosis

Main character	Clinicians	Patients		Total
		No BPD	BPD	
Human	4	11	2	17
Non-human	20	10	17	47
Total	24	21	19	^a 64

^a Three stories came from patients who did not undertake a SCID-II interview and could not therefore be given a definitive DSM-IV Axis II diagnosis. They bring the total number of stories collected in the study to 67.

Patients with no BPD diagnosis were more likely to choose a human main character than either clinicians or patients with a BPD diagnosis. Stories from clinicians and patients with BPD were compared to see whether there were any differences in the proportions of the five categories of main character, and there were not ($\chi^2 = 5.79$, $df = 4$, $p > .05$).

13.0 QUANTITATIVE TEXTUAL ANALYSIS:

DISCUSSION

13.1 DISCUSSION OF STORY LENGTH

From clinical experience it was expected that there would be a positive correlation between better mental health and length of stories. The association was not quite that simple however. It appears that among patients with at least one personality disorder diagnosis, those who left school at or before 16 produced significantly shorter stories.

What is interesting is that patients with a personality disorder diagnosis who continued in education after 16 produced stories just as long and elaborate as those without a personality disorder diagnosis; in fact the distribution of story lengths for the former subgroup was indistinguishable from the story lengths produced by the clinicians in the study. So story length, while an apparently very stable variable from one occasion to another, is not a good predictor of the presence of personality disorder, and probably has more to do with educational history than current diagnostic status.

What could not be investigated here is the relationship between the length of time spent in education and the complexity of personality disorders in adulthood. It is likely there is a causal connection but a plausible argument could be made for the causation going in either direction. Emerging personality problems might make it more likely that a person leaves school early. Conversely, leaving school early might result in more complex personality problems in later life, by increasing socioeconomic stresses.

The partial correlation carried out established that the biggest source of variation between the length of individual stories was the level of education of the author (30%), followed at some distance by the level of personality disturbance (12%). But if the DSM-IV definition of personality disorder as arising in childhood or adolescence is accepted, almost all of the patients now diagnosed with a personality disorder will have begun experiencing its effects while at school. This may well have had an effect on the level of education reached and so the partialling out of this variable may not strictly be necessary. Either way, overall transcript length remains a significant difference between patients with and without a personality disorder.

The ratio of initial story length to answers provided to elaborating questions had also been expected to be different among different groups. Clinical experience had suggested that story authors such as students, trainees and fellow clinicians produced a greater proportion of their total words in the initial telling of the story than did patients. This proved to be the case, suggesting that the clinicians in this study provided more of the detail of their story in its first telling, whereas on the whole patients needed further prompting with the scripted questions before providing this detail. Overall story length is similar between clinicians and patients without a BPD diagnosis, but the two groups are distinguished by the significant differences in the ratio of initial story to question responses.

13.2 COMPUTERISED TEXT ANALYSIS

There are some striking overlaps between the categories identified by the three text analysis programs. For example, all three of General Inquirer (GI), Linguistic Inquiry and Word Count (LIWC) and the Regressive Imagery Dictionary (RID) detected categories to do with relativity of space and time. Words from these categories were significantly more likely to be used in stories from patients with a diagnosis of BPD.

Table 13.1: Dictionary categories describing relativity in time and space

Dictionary	Category label	Description of category	Examples of words
LIWC	SPACE	Words to do with relative positions in space	Around, over, up
RID	Primary process/ Concreteness	Words indicating concrete physical description	Across, behind, distant, flat, near, side, west, wide
GI	TIMESPC	Words relating to space and time	Afternoon, central, elapse, height, instant, outer, past, route, tall.

Patients with a diagnosis of BPD also used words from categories to do with cause, effect and change more frequently.

Table 13.2: Dictionary categories describing cause, effect and change

Dictionary	Category label	Description of category	Examples of words
GI	CAUSAL	Words denoting presumption that occurrence of one phenomenon is necessarily preceded, accompanied or followed by the occurrence of another.	Accident, because, effect, likely, reason, result, why
GI	VARY	Words indicating change without connotation of increase, decrease, beginning or ending.	Alternate, evolve, happen, modify, process, return, shift, variety, waver

While conversely words from categories to do with cognitive processes were more likely to be used in stories by patients without a BPD diagnosis.

Table 13.3: Dictionary categories describing cognitive process

Dictionary	Category label	Description of category	Examples of words
RID	Secondary process/Abstract thought	Words indicating abstract thought processes	Analyse, believe, calculate, decide, guess, real, suppose, weigh
GI	SOLVE	Words (mostly verbs) referring to the mental processes associated with problem solving.	Analyse, consider, guess, reason, select, understand, weigh
GI	KNOW	Words indicating awareness or unawareness, certainty or uncertainty, presence or absence, as well as components of mental classes, concepts or ideas.	Certain, difference, idea, imaginary, logic, similar, suppose, unknown, view

Perhaps the stories from patients with a BPD diagnosis tend to be more concrete and to do with physical action rather than intellectual understanding. This would fit with a model of Borderline Personality that sees BPD as a state where impulsive, physical acting out takes the place of careful, considered thinking. Another group of categories used more frequently in stories from patients with a BPD diagnosis concerns themes of absence and loss.

Table 13.4: Dictionary categories describing absence and loss

Dictionary	Category label	Description of category	Examples of words
RID	Others/Negation	Words indicating negation	Absence, blank, never, nobody, not, nowhere
LIWC	SAD	Words to do with affective processes of sadness, depression	Grief, cry, sad

And this can be contrasted with a category used more frequently by patients without BPD diagnosis:

Table 13.5: Dictionary category describing human roles and relationships

Dictionary	Category label	Description of category	Examples of words
GI	HU	Words referring to humans, their social roles and categories.	Brother, companion, family, grandchild, Mrs, prisoner, role, woman, youngster

It is hardly surprising that negative, grief-laden words are more common in the stories of the BPD group, and that words referring to human and social roles are correspondingly less frequent.

Two further linked categories are used more frequently in stories from patients without a BPD diagnosis. These are words to do with emphasis and certainty:

Table 13.6: Dictionary categories describing emphasis and certainty

Dictionary	Category label	Description of category	Examples of words
GI	OVRST	Words indicating emphasis in realms of speed, frequency, quantity accuracy, size, intensity, certainty and extremity	Abominable, delirium, great, justify, matchless, perpetual.
LIWC	CERTAIN	Words to do with the cognitive process of certainty	Always, never

It is a little surprising to find these words used less often in stories by the BPD group. It might have been expected that the splitting, black and white, all-or-nothing type of thinking characteristic of BPD might have been expressed in these kinds of words. Perhaps the ambivalence of the borderline position means that stories from the borderline group are in fact expressing more doubts and fewer certainties

Overall then, the word use in stories from the BPD group can be contrasted with stories from other patients in the following way:

Categories of words used more frequently in stories from the BPD group:

- Descriptions of physical or temporal location
- Explanations of cause, effect and change
- Words denoting loss, isolation or grief

Categories of words used more rarely in stories from the BPD group:

- Descriptions of rational, cognitive processes
- Words describing humans and their roles
- Words implying emphasis and certainty

The four LIWC categories identified in this study, and the more general categories described above, are not consistent with the LIWC factors identified previously. Pennebaker and King (1999) identified four factors comprising 15 of the 72 possible dimensions of the LIWC. Only one of the four categories identified in this study (the DISCREP category, including words like could, would, should) was among the 15 identified by Pennebaker and King. However their large factor analytic study involved written text, not transcriptions of speech, and they point out that (p.1309) "we are not yet able to generalize the current factors to natural speech". They make the further point that although "the ways people express themselves in words are remarkably reliable across time and situations" (p.1308), the investigation of language use is at an early stage and it is not always possible to make testable predictions from theory

about how language will be used. Instead it is necessary to use observational methods to identify patterns which could then be used for theory building.

Other studies have used text analysis procedures with psychiatric outpatients (Gottschalk, Stein, & Shapiro, 1997) and to compare work by poets who committed suicide with work by non-suicidal poets (Stirman & Pennebaker, 2001). The first study did use natural speech, demonstrating that 5-minute samples of speech could be processed using a content analysis system (the Gottschalk-Gleser Content Analysis Scales). Subsequent analysis demonstrated consistent levels of correlation with various instruments such as the MMPI and SCL-90, although these correlations tended to be low to moderate (the highest was .45 and most were in the range .25 to .35). These levels of correlation are not of immediate clinical use with an individual patient, but may distinguish groups and subgroups adequately. The present study seems to offer a method resulting in a score that says something about the individual, and not just the group to which they belong.

The study of suicidal and non-suicidal poets demonstrated that text analysis procedures (the LIWC in this case) could be used to distinguish text from two matched groups with very different outcomes. The authors conclude that

“...text analysis can be used as a tool for understanding the way that psychological pain, preoccupation with death and self, and association with thought and feeling can be manifested in writing and potentially predict (or indicate the current state of) psychological and emotional health.” (Stirman & Pennebaker, 2001: p521)

Once again the LIWC categories identified as discriminating between the two groups of poets were not ones identified in this study of 6-part stories, demonstrating that there is some way to go still in establishing a gold standard set of categories - if indeed this is ever to be feasible.

In this study the reliability across time of the GI-BPD and LIWC-B scores has been demonstrated. Further studies could use more than two time points to further demonstrate this. The discriminant ability of the LIWC-B score would also need to be tested against further transcriptions from patients with and without a diagnosis of BPD, to rule out the danger of circularity in a tool developed from a group of stories then being tested upon the same group.

One technique used to guard against this is to randomly select half the words from each piece of text, construct the instrument and then test it on the remaining words. This was not done in this study for two reasons. First, in this method the testing phase is not done on entirely separate pieces of text from new subjects, so the danger remains that similarities detected are within-subjects similarities rather than between-subjects similarities. Second, it had been established that shorter amounts of text (the initial story only) were not sufficient to reliably detect patterns among the General Inquirer categories. For this reason it was felt necessary to use the fullest amount of text possible for the construction of the scores, and to leave testing these for a subsequent study when new text from new patients was at hand.

The failure of this study to replicate the findings of other studies using the LIWC suggests that generalisation to wider groups would be unwise. The LIWC

categories identifying suicidal poets, and the LIWC factors found from a large student population were not found here. The four LIWC categories proposed as a scale have been derived from a group of adult out-patients of community mental health services. The only thing the scale purports to identify is the presence of Borderline Personality Disorder. Generalisation to other groups or other disorders would need further work to be undertaken. The fact that the LIWC-B scale identified a third of the clinician stories as scoring over the threshold for BPD reinforces the caution. Either the clinicians in the study have an unusually high level of personality pathology among them, the scale's discriminant ability is an artefact, or the scale should only be applied to stories produced by patient subjects and not more widely. The first possibility is unlikely, the second suggests the need for a confirmatory study, and the third would indicate a precautionary approach in applying the technique only to groups similar to those in this study.

It is not being suggested that the LIWC-B score, or any other to be derived in this way, might become a part of routine clinical practice. The taping and transcription takes up a great deal of time and the analysis needs to be done with a computer program that would not be readily available outside a research group. Rather, the validity and reliability of the GI-BPD and LIWC-B scores is called as evidence that 6-part stories can be investigated systematically and be shown to have stable properties from which clinical inferences can be drawn.

13.3 CUSTOM-BUILT DICTIONARY FOR BASICPh

It was disappointing that the six BASICPh categories showed no differences between the BPD group and the rest of the patient group. This is particularly so when considering some of the categories of words found to distinguish the BPD group from others:

Categories of words used more frequently in stories from the BPD group:

- Descriptions of physical or temporal location - cf BASICPh Physical category
- Words denoting loss, isolation or grief - cf BASICPh Affect category

Categories of words used more rarely in stories from the BPD group:

- Descriptions of rational, cognitive processes - cf BASICPh Cognition category
- Words describing humans and their roles - cf BASICPh Social category

There may be several reasons why this was the case. First, the BASICPh system itself has a theoretical and clinical provenance (Lahad, 1992; Lahad & Ayalon, 1993) but no empirical base. Perhaps it is simply based on faulty theory and therefore is unlikely to be demonstrated empirically. Or possibly the BASICPh distinctions really do exist, but have no bearing on personality disorder. If we had recorded subjects' height as one part of the dataset, it would be unlikely to correspond in any way with the clinical data, although it would be a replicable, externally verifiable variable itself. A third possibility is that the dictionary was inadequately assembled. The only external sources used were

the clinical examples provided in the chapters by Lahad cited above, along with the set of synonyms and inflected forms found in the dictionary building tool of WordStat. Classical dictionary-building usually involves external verification by expert panels to confirm the face validity of the categories. This was not undertaken here because of the very large demands it would have made on external raters for a relatively small part of the overall study.

The text corpus for this study remains in place, will be kept and can be added to in the future. With a larger body of transcripts from a wider range of subjects, further investigation would be possible. If the BASICPh categories do not demonstrate potential for further investigation, there may be other theoretical or evidence bases from which other custom built dictionaries could be built to greater effect.

13.4 OTHER FEATURES OF QUANTITATIVE TEXT ANALYSIS

The preponderant choice of male main characters by both male and female subjects is striking. It could be argued that women, in choosing male main characters, are simply obeying the instructions in the protocol for eliciting the story. These include the instruction that participants should “make this main character someone or something as unlike you as possible.” Perhaps women are simply better than men at following this instruction and creating a main character of the opposite sex to themselves. This seems unlikely, and a simpler explanation would be that women, like men, are exposed to a body of stories that overwhelmingly have men in the lead role. Only four Shakespeare plays have women in the title and two of these are the *Merry Wives of Windsor* and

the Taming of the Shrew! (By contrast there are 25 play titles with men in them.) Novels, films, TV soaps, non-fictional stories such as news or sports reports; it is hard to think of any category of story in which female protagonists outnumber males. The only category (and this an arguable one) where this might be the case is the genre of romantic fiction.

These results may provide some evidence of the projective process at work. Only one of the 19 male subjects chose a non-male main character, suggesting that despite the instruction to “make this main character someone or something as unlike you as possible” male subjects chose main characters of the same gender as themselves. This could have been entirely due to societal and cultural expectations about the gender of leading characters in stories. However despite this conditioning and the instruction given, eight of the 48 stories from female authors had a female main character. The projective impulse was in these cases strong enough to override the general cultural expectations and the specific instructions of the 6PSM.

The differences in the type of main character chosen are interesting. The contingency table below has already been shown:

Table 13.7: Human and non-human main characters chosen, by author group

Group	Type of main character		Total
	Human	Non-human	
Clinicians	4	20	24
Patients			
No BPD	11	10	21
BPD	2	17	19
Total	17	47	64

Since a chi-squared test ($p < .01$) suggests that this has not occurred by chance, how are we to make sense of this? Post-hoc partitioning of the table

suggests that the most important factor is the surprisingly large number of human main characters chosen by patients without a BPD diagnosis. It might be imagined that there would be a kind of continuum from clinicians to patients without a BPD diagnosis and ending in patients with BPD, but this is not the case. The two extreme groups (clinicians and patients with BPD) resemble one another in having relatively few (6 out of 43) human main characters while the median group of patients with no BPD have a majority (11 out of 21) stories with human main characters.

The only detectable difference in the choice of main character between these two groups was that the clinicians seemed to be more likely than the patients with BPD to choose a fantasy main character, whether animal or human ($\chi^2 = 4.05, df = 1, p < .05$).

So perhaps if there is a gradient in the choice of main characters consistent with the gradient of pathology described above, it is this:

Fantasy characters (human or non-human) > Realistic human characters > Realistic animal characters

This is borne out by the contingency table below, which confirms that there are significant differences along these lines (chi-squared=11.45, df=4, p<.05).

Stories with inanimate main characters were excluded from this analysis, as they were small in number and spread evenly among all three groups.

Table 13.7: Fantasy and realistic main characters chosen, by author group

Group	Type of main character			Total
	Fantasy (human or other)	Realistic human	Realistic animal	
Clinicians	9	4	8	21
Patients				
No BPD	4	11	6	21
BPD	2	2	9	13
Total	15	17	23	55

These differences are not overwhelming, and in the clinical situation one would not want to base a diagnosis on something as tenuous as the choice of main character in a 6-part story. But this is a potentially fruitful trend and if repeated in a larger study it would be interesting to try and evolve a hypothesis about why these differences arise.

14.0 QUALITATIVE ANALYSIS: METHOD AND RESULTS

14.1 FOREWORD TO THIS CHAPTER

In line with common practice among qualitative researchers, I have written this chapter in the first person and with a more deliberately self-conscious 'voice'. The main reason for this is to acknowledge the deep personal investment necessary in all stages of this Grounded Theory building, but particularly the later ones. As concepts evolve into categories and a theory begins to take shape, the hand of the qualitative researcher becomes increasingly evident. In quantitative analysis there are devices such as double blinding and the testing of inter-rater reliability that can, to some degree, remove the analyst from the picture; in this case a third-person account is appropriate. For this portion of the analysis I have chosen to emerge from the fiction of the third person, in order that my own biases and preconceptions can be acknowledged, owned and (if only partially) thereby accounted for. For example, at the beginning of the Results section of this chapter I write "The categories that I identified were as follows:", which originally I had drafted as "The following categories emerged:". The difference in the subject and verb of the two sentences is important; these categories did not emerge by themselves under their own steam, they were consciously assembled by me. Writing in the first person makes my active presence more obvious and acknowledges that these categories are my own constructions, not necessarily naturally occurring phenomena that I have uncovered.

It is also customary for the qualitative researcher to declare the position from which s/he comes. Mine is that I am a dramatherapist by profession, with an interest in the 6-Part Story Method that goes back about a decade. I have a

strong interest in creative, spontaneous ways of working and I have been influenced by authors such as Jung and Winnicott, but I have retained a core of scepticism about the unprovability of some psychoanalytic concepts. Unusually for a dramatherapist, I am more at home with quantitative than with qualitative research methods, so I approached this section of the analysis with some caution.

14.2 AIMS OF THE QUALITATIVE ANALYSIS

In addition to data I collected for quantitative analysis, I also collected some qualitative data in the form of participants' responses to the storymaking process. The aim of this was twofold:

1. To establish the degree of acceptability (or otherwise) of the 6PSM as a method of assessment and to canvass suggestions for amendments to the method.
2. To provide raw material for the development of a Grounded Theory (Glaser & Strauss, 1967; Strauss & Corbin, 1998) to begin to explain the action of the 6PSM.

14.3 METHOD

14.3.1 DATA COLLECTION AND PREPARATION

I used two methods to elicit participants' reactions to the process. Firstly, the final questions on the protocol for tape recording 6-part stories departed from

questions about the story itself and asked about the storymaking process. All 67 story tapes therefore had some responses to questions such as:

“What’s it been like doing this storymaking exercise, what are your first reactions to it?”

and

“What do you think is happening when you tell the story and we discuss it?”

Secondly, I taught clinicians the 6PSM method in groups of two or three. At the end of each teaching session, after their stories had all been individually recorded and discussed following the strict protocol described above, I recorded a further group discussion. This was not formally convened as a focus group (Morgan, 1997) but had some of the features of such groups. I acted as facilitator and asked open questions aimed at stimulating discussion between the participants reflecting on the process they had all just been through.

The responses gained by both methods were transcribed by me and combined into two documents, with speakers individually identified by code numbers. I then imported the documents into the NUD*IST content analysis program (Richards & Richards, 1991) for qualitative analysis. (NUD*IST stands for Non-numeric Unstructured Data, Index Searching and Theorising.) Between them the documents contained 10,400 words of transcribed text from all 49 participants (24 clinicians and 25 patients.)

14.3.2 DATA ANALYSIS

The method of analysis I used was based on the Grounded Theory first developed by Glaser & Strauss (1967). The detailed procedure I followed was that described by Strauss & Corbin(1998). It was not a classic, full Grounded Theory study as described by Strauss & Corbin as it lacked some of the depth and rigour this would require. In a full study for example, questions would not be scripted and asked in the same way of every participant, but as part of a more flexible conversation between participant and researcher. However most of the text I analysed had been gathered by the clinicians following a fixed script, from which they were under instructions not to deviate.

Another difference was that all the data were collected before analysis began; in a formal Grounded Theory study early results would influence the questions to be asked of later participants so that hypotheses could be generated and tested as the study progressed. Emerging theories might be tested by further theoretical sampling to target a group of interest.

A third difference arose from the decision to carry out a mixed methodology (qualitative and quantitative) study. Planning for the quantitative elements of the study meant that I carried out some preliminary literature searching and hypothesis generation early in the process, much sooner that would usually be the case in a pure qualitative study.

Some key features of Grounded Theory were retained however. The first was that of constant comparison; this means that throughout the process of analysis

the raw text data were retained as a primary check on any more elaborate theorising which arose from it. For example, if I developed a tentative theory in response to a particular piece of text, then I examined the rest of the text to see whether that theory was supported by the data or not. In the latter case the observation was not discarded, but rather was retained as an exception needing explanation or a limit case defining the boundaries of a phenomenon.

The second feature was that of data saturation; this relates to the problem in qualitative analysis of knowing when to stop gathering data. Being unable to use quantitative methods of sample size selection, the idea is that data are collected and coded (see below for a description of coding) until no new codes emerge. At this point saturation is said to have been reached and data collection can stop. In this study I stopped data collection when all participants had been tape recorded, but the principle of saturation was still observed. I coded text progressively and towards the end of coding the number of new codes being added dwindled to nothing, while old codes were simply being added to with repetitive, similar data. I took this to demonstrate that saturation had been achieved and that nothing would have been gained by recruiting and interviewing more participants.

Thirdly, I adhered to the process of open and axial coding to produce concepts and categories, each with their own properties and dimensions, as far as possible. This coding process will now be described in more detail.

14.3.3 CODING

14.3.3.1 Identification of concepts

I read through entire text of the transcripts under consideration, looking for the presence of concepts that might be of interest to the central question "How do participants experience the 6PSM process?" When a concept was found in the text I coded it and gave it a name. The concept might be embodied in a single word, but the whole text-unit containing the word would be coded. For the purposes of this analysis, each separate line of text in the NUD*IST display was treated as one text-unit. This comprised a maximum of 70 characters (including spaces). A single text-unit might be coded under several concepts. The coding process involves the NUD*IST program invisibly marking that text-unit so that it could be retrieved later. I gave concepts names that were as memorable and illustrative of the concept as possible, so that later on in the process if the same concept were illustrated a second time in a subsequent text-unit it could be marked with the previously established code.

As an example of this, the following eight text-units from the interview with Clinician 20 are shown along with their codes:

TEXT-UNIT	CODES
1 ...it felt OK at the end even though I did feel a bit	A
2 exposed I suppose how much it relates to you, that's	B, C
3 what I think is surprising. You don't realise it's about you	D, E
4 when you begin, but it's you, you can actually think it's	C
5 about cats and what a lovely life, but even within that	F
6 context and what a lovely life it can be quite difficult.	G
7 And that's what I think is so empowering, it's like	H
8 that's me in there somewhere, bits of it anyway.	C

- A Calm at end
- B Exposing, transparent
- C Sees self in story
- D Surprising elements
- E Started trivially
- F Seems irrelevant at first
- G Process is difficult
- H Empowering

By the end of the first pass of the coding process, approximately 2,100 separate codes had been applied to text-units, comprising about 75 different concepts. In practice I made several passes, in case categories that only emerged towards the end of the analysis might be useful to code earlier text-units.

In addition, I made use of some of the features of NUD*IST to make coding easier. One of these was the string search, which searches all the text under study for instances of a particular word string. For example, it became clear that one important concept was that of surprise, where participants described results or experiences they had not expected. In case I had missed coding examples of this concept, I performed a string search for the string 'surpris*' which would pick up any string beginning with those letters. Six text-units were identified, only four of which had already been coded, so two further pieces of code could be added.

14.3.3.2 Amalgamation of concepts into categories

The 75 initial concepts (referred to in NUD*IST as nodes) were initially unstructured and in no relationship to one another, what NUD*IST describes as free nodes. The next step was to gather these concepts into categories on the basis of their similarities with or differences to one another. For example, concepts A, B, G and H above were gathered into a category labelled 'Affective reactions to process'. This category was further divided into positive reactions such as A and H, negative reactions such as B and G, and neutral reactions. Some concepts were allocated to more than one category, for example categories A, E and F were also allocated to a category labelled 'Start and end of process contrasted'.

NUD*IST was helpful in constructing categories as well, by taking free nodes (concepts) and attaching them to one another to form the start of a hierarchy. When several concepts had been clustered together it became clear that I was now dealing with a category, which could be labelled as such.

However the most flexible way of dealing with the 75 concepts was to copy the list of free nodes from NUD*IST, print them out, cut them into strips and spread them out on a large table. This allowed groups of concepts to be tentatively assembled, as well as allowing some concepts to straddle groups and giving a sense of which categories were closely related to one another and which were more distant.

14.3.3.3 Finding properties and dimensions of categories

Once the categories had been established, their properties and dimensions were investigated. These are defined thus:

To further clarify, whereas properties are the general or specific attributes of a category, dimensions represent the location of a property along a continuum or range. (Strauss & Corbin, 1998: pp117, emphasis in original)

Strauss and Corbin give the example of a notional category called 'Flight', which might have one property of 'speed' ranging along a continuum from slow to fast, and another property of 'height' from low-level to high. Individual instances (concepts as illustrated in text-units) will all fall somewhere along each of the continua identified.

In this study some categories could be elaborated in these terms, whereas others could not. This was probably because the interview texts being analysed were almost all responses to a fixed, scripted set of questions rather than being part of a free flowing, developing conversation. The latter might have given more opportunity for properties and dimensions to emerge naturally.

14.3.3.4 Axial coding

This process involves inspecting the categories that have been identified and seeing how they may relate to one another, usually by means of finding linking properties and dimensions. According to Strauss and Corbin:

When analysts code axially, they look for answers to questions such as why or how come, where, when, how, and with what results, and in so doing they uncover relationships among categories. (1998: pp127)

It is recognised that in doing this the analyst is now working with abstractions that s/he has created, and necessarily moving away from the raw data of the words in the text-units. This is particularly the case in the last stage of axial coding, the selection of the central category (sometimes called the core category.) I will discuss this further when reporting on the results obtained later in this chapter, but essentially the central category must 'pull the other categories together to form an explanatory whole.' (Strauss & Corbin, 1998: pp146). When I was at the planning stage, I hoped that this central category might go some way towards answering the question "How does the 6-Part Story Method work?"

14.4 RESULTS

14.4.1 IDENTIFICATION OF CATEGORIES

The 6 categories and 3 subcategories that I identified are listed below, along with those properties and dimensions that were immediately apparent. These will subsequently be described in more detail. The categories are listed in order of the number of text units coded to them.

Table 14.1: Categories identified through free coding, with their Properties and Dimensions

Category & Subcategories	Properties	Dimensions
Parallels between story and own situation	Relevance	Irrelevant-relevant
Progressive development of story	Time	Beginning-end
Response to process	Evaluation Strength	Negative-Positive Weak-Strong
Comparisons between first and second stories	Similarity Difficulty	Similar - Dissimilar Becoming harder-becoming easier
Aesthetic distancing	Distance Effectiveness	Nearer-further Less-more
Material evoked Affective Cognitive, factual 'Subconscious'		

14.4.1.1 First category: *Parallels between story and own situation*

Altogether there were 1529 text units in the transcripts analysed. There were 487 text units coded to the category Parallels between story and own situation.

The majority (414 text units) indicated that participants thought the story was a good metaphor for their own life and situation:

Yes, that's what happens within my family circle. That's too freaky. You're getting all this just from pictures, I'm just talking about these few pictures that I've drawn and now I can really see how this story relates to different aspects of my life. (Patient B11, borderline diagnosis)

... you're definitely seeing what a person's outlook is. It can measure how depressed a person is and what his approach to life is. (Patient M8, no BPD)

I thought it would just be a paper exercise initially, but what I found is that the blue cat is a little like me, I do become bored easily, I do have some good friends who don't look for things in return. (Clinician C20).

A smaller number of text units (73) indicated that participants thought the story was not relevant in this way:

I think it's quite unlike me actually. No, I think it's quite unlike me. (Clinician C18)

(Responding to question "Can you see bits of yourself in there anywhere?") No, not really because I think you were asking to try and push it away as far as possible from me at the beginning, so I don't think it does, I had that in mind all the while. I suppose I might see me with the

struggles I suppose, but I suppose that is everybody we all struggle with something. But I don't think so, no. (Patient M3, no BPD)

Interestingly, the same patient had a similar reaction to telling the story the second time:

I don't think so, I don't think so. I'm generally a good person. Maybe the Joe Bloggs could be represented by me but I don't get offered this box, I might open it. (Patient M3, no BPD)

But even more interesting were his general comments about the process:

Very difficult, especially not being able to use words. Yeah, a bit nervy I suppose it is like when it is something new, don't explore these sides very often...All in all I find it quite comfortable to do, but there is little pieces where I feel very insecure about it, they're very strange new things, very strange. (Patient M3, no BPD)

Given that he did not think the story showed anything relevant about himself, what sides of himself did he think he was exploring? And what were these very strange new things he was feeling insecure about? Unfortunately the fixed format of the interviews carried out by clinicians did not permit exploration of these paradoxes. However these comments may suggest that even participants who asserted that the story did not tell anything about them may in fact have been defending against uncomfortable truths.

14.4.1.2 Second category: Progressive development of story

The next largest category (272 text units) described Progressive development of story, where participants drew attention to changes that occurred as the story session progressed:

Some were immersed in the storymaking process, only later becoming aware of its possible significance:

I think trying to unravel subconsciously what is happening in my head, going round and round in my head. Although when I am writing the story it doesn't seem that way, it is only afterwards when I'm telling the story that it seems to. (Patient B10, borderline diagnosis)

While others described an initial reluctance or doubt:

A bit dubious at first, a bit cynical at first. I didn't think I could draw the story. But I did and I got to the end. (Clinician C16)

And others described strong emotional reactions at the start which by implication were modified or overcome later in the process:

It felt daft at first, and the trouble is that when I get asked questions I just blank off. (Patient B5, borderline diagnosis)

Within this category there were a number of responses indicating that at the start of the process the story seemed puzzling or inconsequential, and that there was no initial plan of where the story would end.

Interesting because it starts off at not very much and then ends up kind of going, spreading out into all sorts of things and you can see yourself trapping yourself into themes from your own daily life I guess. (Clinician C13)

I didn't know where I was going when I started. (Clinician C3)

I sort of started with a character and I hadn't really given it much thought what I was going to do, I just started with a main character and then the plot evolved as I went on picture by picture. (Patient M7, no BPD)

14.4.1.3 Third category: Response to process

This category was organised strongly across two dimensions, the first being evaluation from positive to negative. There were 135 text-units containing negative evaluations, such as:

What's bad about it is that I'm a crap drawer. My drawing looks like a three or four-year-old. I can't draw. That's what's been bad about it. (Patient M12, no BPD)

I found it quite difficult, I found that I thought I was making up a bit of a silly story really. (Clinician C18)

This evaluation was from the same participant reported earlier who had said:

I think it's quite unlike me actually. No, I think it's quite unlike me.

(Clinician C18)

When looking at the transcript as a whole, this seemed to be an unequivocally negative experience for that particular participant. Other negative evaluations were more ambiguous however, or were in the context of more positive comments:

I think it's a bit scary really because you've got to be very careful about what's safe and what isn't, because it always ends up being a lot bigger than the pictures that you first drew. (Clinician C14)

It's quite stressful, quite nice doing this story, I got into it a bit. To be honest it feels a bit worrying, what am I saying about myself telling this story. (Patient B8, borderline diagnosis)

These two extracts were coded as negative evaluations because of the presence of concepts such as anxiety, lack of safety, stress and worry.

However in both these cases and several others it was the actual (or feared) effectiveness of the storymaking process that was causing the stress. The ambiguity is well illustrated in the oxymoronic "...quite stressful, quite nice..." comment above.

At the other end of this dimension were positive evaluations such as:

I found it quite easy, I enjoyed it. I like doodling and colouring anyway so it was something that was particularly fun to me. (Clinician C1)

It was quite fun really. I've not done anything like this in a long time. It's quite relaxing. (Patient B11, borderline diagnosis)

Interestingly, even participants with strongly negative, anxious, worried reactions (as well as those with positive reactions) tended to report that the stories they had produced were good and relevant pictures of themselves.

The other dimension within this category was that of strength of response. All the responses listed so far could be seen as strong, but there was a smaller number (39 text-units) of responses that were minimal or neutral, such as:

It was ok. Don't know, haven't really got any reaction to it. (Clinician C15)

Not a lot to say really. (Patient B2, borderline diagnosis)

Not sure on that one, my mind's blank. (Patient B10, borderline diagnosis)

None of the participants responding in this way described the stories they had produced as telling anything significant about themselves.

14.4.1.4 Fourth category: Comparisons between first and second stories

This category had 183 text units associated with it. These came from the 18 patients who recorded two stories and were asked, on completing the second story, to compare the two experiences. Many comments were about which session was easier, but there was no clear pattern as to who found the second session easier or harder.

I think I found it more difficult last time just to, when you make up the character that's deliberately very different from yourself, to then have to follow the story through it I found to a bit difficult. (Patient M1, no BPD)

It is harder doing it [the second time], because I felt like I should know what the story was going to be about. (Patient B8, borderline diagnosis)

A few commented on whether there were similarities or differences in the themes. One person pointed out that the similarities might be an artefact of the procedure:

I tried to make it a totally different story. I suppose there are slight similarities where there's a task to be done, there's somebody opposing it and somebody helping it. There are certain similarities to the story because of the instructions of the story to follow. (Patient M7, no BPD)

14.4.1.5 Fifth category: Aesthetic distancing

The penultimate category needs some further explanation. This is a concept well known to dramatherapists, and has been described as one of the core processes for the discipline (Jones, 1993). It describes the phenomenon where the effect of a dramatic role upon an audience depends not on the role's closeness to the audience but on its distance. Paradoxically, the less commonplace and more universal the themes being played, the more relevant they are felt to be.

Aesthetic distance, the structure of the dramatic event, is responsible for both identification and universalisation, idealising the first while humanising the second. (Andersen-Warren & Grainger, 2000: pp14)

This use of a category imported from another source in the literature is not encouraged by authors such as Strauss & Corbin. However other authors on qualitative methods suggest that it is perfectly proper to build on the scholarship of others, providing that due acknowledgement is made:

But undoubtedly the academic literature may prove one of the most useful sources of analytic strategy... because previous research or scholarship may have examined issues in a thorough and systematic way. (Dey, 1993: p67)

Additionally, the inclusion of aesthetic distancing was not decided on before looking for instances to 'prove' its validity. Rather the interview transcripts provided examples of a phenomenon for which there seemed to be a pre-

existing model. The category has therefore been retained, despite the fact that this departs from a pure Grounded Theory approach.

In this study the kind of concepts that were gathered into this category included those illustrated by text-units such as:

Yes, even though I tried to be completely free and find a man called Abraham, it has somehow managed to be about me. (Patient M4, no BPD)

I think initially when you're drawing it, it is very much distant, it is very much not you and it's not about you, it's just a story. It's the explaining it, that you add things that maybe you didn't actually think about when you were drawing the picture. You add your own little bits and pieces that come from your own experience and it helps the story to grow. (Clinician C16)

I think there was a gap between drawing the story which had some kind of interesting childish theme, then when you are asked about it you re-interpret the pictures and they become something else, so really the original story that I drew and the way I interpreted the pictures came through my own sort of experience and the way I would look at things. (Clinician 13)

Participants seemed to be saying that the choice of a main character unlike themselves was not an obstacle to telling a meaningful story. Some went further

and suggested that the aesthetic distance was a positive advantage because it allowed for more freedom and spontaneity:

I don't think you can make it a real person, it's good that you have to make it as far away as possible. Because I did try not to think about it but obviously when I first did it I was thinking about the character but the way that it actually asks you, you can't think of anything before this exercise, or I don't think you can. (Patient M9, no BPD)

14.4.1.6 Sixth category: Material evoked

The final major category identified was that of Material evoked, with its three subcategories of Affective, Cognitive/factual and Subconscious. These referred to the three domains of response that participants described as being evoked.

The Affective domain was the most frequent (32 text-units):

I felt a lot of emotion telling my story (Clinician C12)

Initially it's an uncomfortable feeling of doing it and the drawing stuff I was like no, give me a filing cabinet or a phone and I'm OK but don't tell me to express myself. Now it's over I'm OK. (Patient B9, borderline diagnosis)

However there were almost as many mentions of thoughts, facts or memories being evoked:

I'm not very good at drawing, I'm laughing at my own pictures. But it does make you sit and think. (Patient M12, no BPD)

I found it's OK, it's made me think a little bit. It's made me think. (Patient M10, no BPD)

There were perhaps surprisingly few mentions of bringing material unto consciousness; in fact in all the 10,400 words of transcript the word *unconscious* was never used by anyone, and the word *subconscious* only three times:

I think trying to unravel subconsciously what is happening in my head, going round and round in my head. (Patient B10, borderline diagnosis)

... it's up to you to decide what bit of that encoded information you pass on to the person who's trying to extract the information from you. And I think you do that subconsciously anyway. (Clinician C14)

And that's why it think the first [story is] better and I think I could relate better to that one because I wasn't thinking about what I was doing, and it was more things which were probably going round in my subconscious. Whereas this is a conscious piece of work that I've just done because I knew what I was doing, whereas the first one wasn't, so it was more about me because it was just something that came out. (Patient B11, borderline diagnosis)

Of course if the 6PSM is really allowing a person to communicate unconscious material, they may simply be unaware of the fact that they are doing so. One participant, although not using the word unconscious, did consider this a possibility:

I think it's a really good idea and I think it's very useful, irrelevant of how the person is doing it understands whether they have got a problem or not I do think that it will show a lot of the inner feelings of somebody, whether they were aware of what they were actually doing it or whether they were just doing it because they were told to do it. It would give good insight. (Patient B7, borderline diagnosis)

14.4.2 AXIAL CODING

The categories were examined in detail for properties and dimensions that went beyond the initial few identified in Table 10.1, in order to try and find links between the categories. It became clear that one property, that of time, could actually be applied to all categories. A great many of the participant transcripts were in the form “at first I experienced X, but then later on I experienced Y”. They described a low intensity experience becoming heightened, a high intensity experience becoming lowered, or one experience being replaced by another over time. The period of time mentioned covered within-session changes and (for those patients who completed two stories) between-session changes. The changes described included these transitions, which do not form an exhaustive list:

Moving (sometimes unintentionally) from unrecognisable to recognisable personal themes:

Even if you didn't intend to go down that line it just seems to appear, and then you start recognising bits of yourself and then bits of your life that happen, even if you don't intend it to. (Clinician C11)

And as I've been saying it -and as I was doing it actually, when we got to about the fourth picture I started thinking hmmm, you know, even I could start to make parallels between what I've drawn and me. (Patient M9, no BPD)

Moving from evoking less emotion to evoking more emotion:

Moving from not knowing where the story was going to knowing where it has been:

I think I found it quite emotive, not so much when I was drawing it, and I didn't know where I was going when I started. But yeah, I felt it was powerful, a powerful way of expressing some things that maybe I needed to. (Clinician C3)

Moving from blankness to being able to think and create:

Moving from doubt to satisfaction:

I was surprised I could think anything at all, I didn't think I could at the beginning. But now I can go on to do things better in your way. I was surprised because I didn't think I would be able to do anything at all when I first started, but I feel satisfied. (Patient M6, no BPD)

Moving from apprehension through upset to OK:

I was apprehensive about doing this because of that feeling of being exposed and trying to keep it simple, but it was upsetting when you were looking at the things that made it difficult, but that for me was counterbalanced by the things that made it easier because I was able to identify with those, so it felt OK at the end even though I did feel a bit exposed. (Clinician C17)

14.4.3 IDENTIFYING THE CENTRAL CATEGORY

This led to me deciding to adopt the category labelled Progressive development of story as the central category. On returning to the transcripts I looked again for any instances of categories where the property of time was important, and added these to the central category. Progressive development of story was then investigated to see what properties and dimensions it might have. These included:

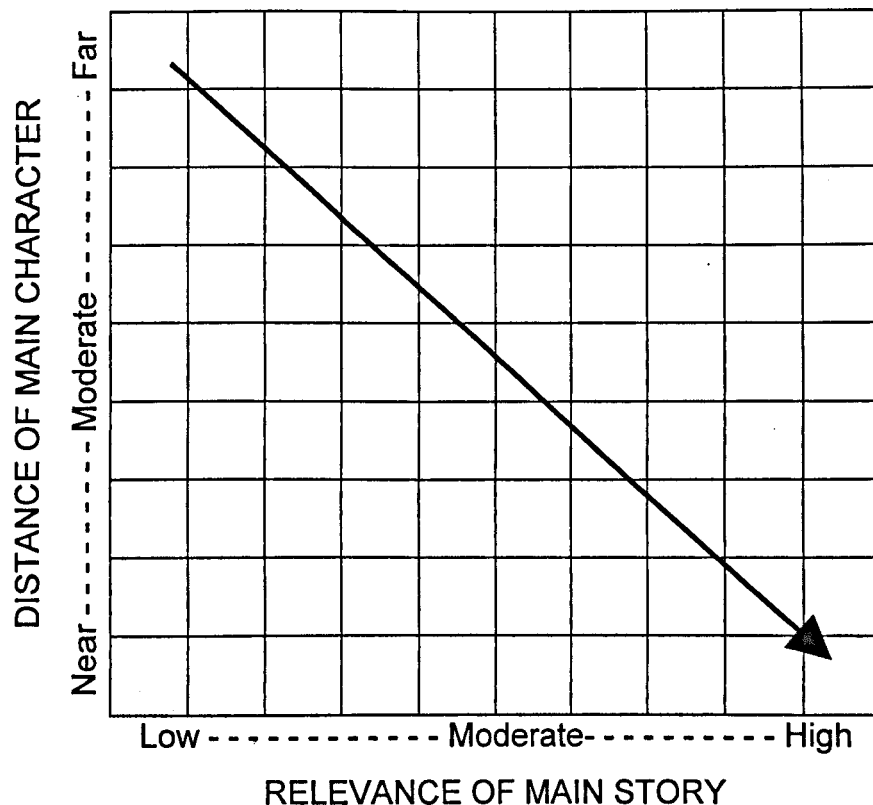
Table 14.2: Properties and dimensions of the central category: Progressive development of story

Properties	Dimensions
Time	Early in story-late in story First story-second story
Quality of emotional experience	Anxious-upset-calm
Relevance of story to own situation	Low-High
Familiarity with instructions and process	Unfamiliar-familiar
Distance of main character and situation	Near-far
Knowing what to do	Not knowing-knowing

These dimensions were then cross-cut with one another and with those of other categories to see what patterns might emerge about the way properties co-varied across two dimensions. For example following the procedure outlined by

Strauss & Corbin (1998: p141), Relevance of story was initially crosscut with

Distance of main character thus:



When individual text-units were mapped on to this grid, it quickly became clear that as the storymaking session went on, most participants moved from a distant main character to one who was seemed closer to themselves, while the relevance of the story moved from low to high. This finding seemed almost trivial, but it was the exceptions that proved more interesting; these were of two sorts.

1. There were some participants for whom there was no movement; the distant main character remained distant and the story's relevance remained low.

2. Other participants reported little movement because there was no surprise in the story. Either a main character was too obviously a metaphor for the teller, or the teller was too self-conscious about the storytelling process. The main character tended to be too similar to the teller at the outset of the story and there was no deepening of relevance as the story progressed.

Both groups described the following experiences:

- Minimal, low-intensity response to 6PSM
- Material evoked mainly trivial, factual
- No development over time
- Minimal expression of early anxiety or late relief

It appeared that tellers in the first group were probably initially dubious (or defensive) about the 6PSM as a method. The doubtful participants probably had their doubts confirmed, while the defensive participants were able to use the initial distancing effect of the method to their advantage by not allowing the identification with the main character to occur.

Tellers in the second group seemed more likely to want the method to work, but were too anxious or self-conscious for this to happen with a natural flow. They were perhaps always too aware of the artifice and the structure, and there was insufficient spontaneity to create anything new and surprising.

However responses of these two sorts were relatively few; most people described these experiences:

- High intensity response to 6PSM
- Emotional material evoked
- Story and teller's response changes over time
- Early anxiety replaced by late relief or satisfaction

For these people, the development of the story took this form.

1. At the start of the process:

Feeling anxious, uncertain, or blank. Wondering what to do. A main character is chosen who seems a long way from reality, comes out of nowhere. The story does not seem relevant.

2. In the middle of telling the story:

Feeling upset or emotional. Story takes on a surprising life of its own. Teller's own issues emerge in those of the main character who becomes more relevant.

3. After telling story and questioning:

Feeling calmer, pleased. Surprised to have achieved so much so quickly. Story seems very personally relevant.

The contrast between those participants for whom the 6PSM works, and those for whom it does not, can perhaps be summed up in the following schematic figures:

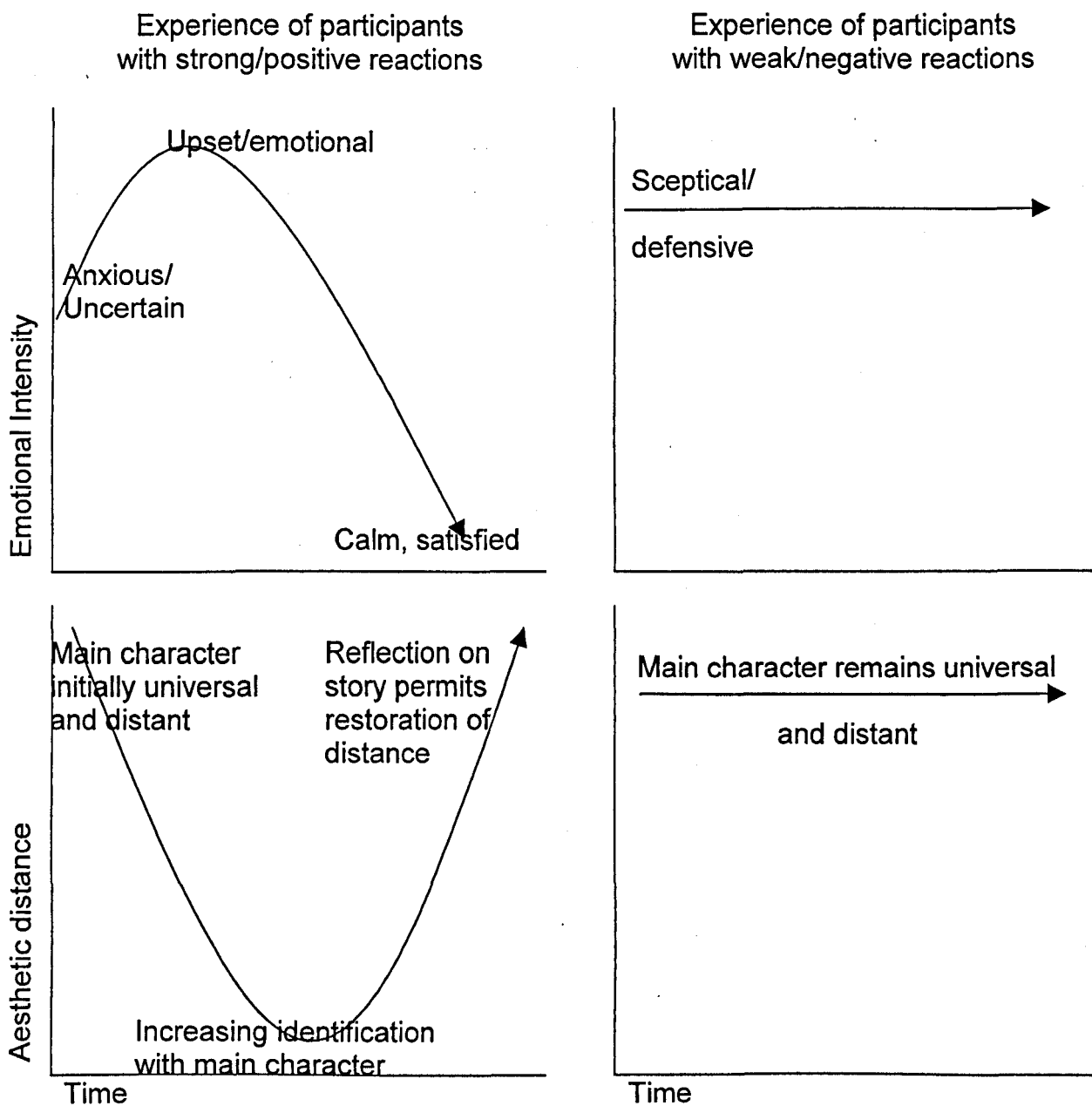


Figure 14.1: Changes in emotional intensity and aesthetic distance over time experienced by participants with positive and negative reactions

Visually, the difference between those stories that seemed to have an impact and those that did not is striking. On those occasions when the story works, the emotional intensity and aesthetic distance are changing constantly, first in one direction and then the other. On those occasions when the process is unsatisfactory there is no such change.

Once this pattern became clearer, I wondered whether the three groups of participants showed any differences in how relevant they had found the 6PSM experience. I looked at the transcript again and coded each participant's reaction to the question about personal relevance of the story ("Now that you've done the story, do you think it tells us anything about you? Are bits of you in there somewhere?") I coded responses as Strongly Relevant, Weakly Relevant or Irrelevant. When tabulated, participants' responses were:

Table 14.3: Participants' assessments of story relevance, by group

Assessment	Group		
	Clinicians	Mainstream CMHT patients	Patients with Borderline diagnosis
Strongly Relevant	13	4	4
Weakly Relevant	1	5	2
Irrelevant	3	2	4

With such a crude, subjective coding on my part, unvalidated by any reference to another rater, it did not feel appropriate to conduct a statistical test of the significance of these results. But it did appear as though clinicians were much more willing or able to see personal relevance in the stories they produced. Conversely, patients with a diagnosis of BPD seemed to be more likely than other participants to reject or deny any personal relevance in the story.

15.0 QUALITATIVE ANALYSIS: DISCUSSION

15.1 ACCEPTABILITY OF THE 6-PART STORY METHOD

This study has not attempted any quantitative survey of the acceptability of the 6PSM, but the preponderance of favourable over unfavourable comments elicited suggests that most participants found it an acceptable method. Those who made negative comments were neutral, dismissive or mild in their statements, rather than appearing angry or distressed. Of course there are some major problems in making this assertion and generalising from this research group to a wider clinical population:

1. Clinician participants came forward as active volunteers, either because they already knew the researcher and his work on the 6PSM or because the description of the research which was circulated to them interested them. They therefore form a group that is predisposed to look favourably on the 6PSM.
2. Patient participants were specifically approached, rather than volunteering themselves, but all were given a brief outline of the 6PSM procedure and the opportunity to decline to take part. Those who might have been less likely to find the 6PSM acceptable presumably took that opportunity to decline to participate, again skewing the sample.
3. Both groups of participants were asked the questions about acceptability by someone who was not necessarily impartial. Clinicians were asked by the researcher about a technique in which they knew he had an investment. Patients were asked by their CMHT key worker whom they might have an interest in pleasing or placating. Critical responses may therefore have been muted or censored for reasons of social desirability.

In the case of the second problem, it should be emphasised that in clinical practice, the 6PSM is not (and could not be) imposed on an unwilling patient. In this sense the research group of patient participants is no different from a general clinical group, because only those patients who are at least willing to give the 6PSM the benefit of the doubt will be likely to agree to undertake it.

With the third problem, it is at least possible to judge the reaction of the participant to the 6PSM through their non-verbal, as well as their verbal responses. No clinician reported either that any of their patient participants had become distressed when undertaking the 6PSM, nor that anyone who had started the 6PSM process asked for it to be halted before completing a story and all the associated answers. On the other hand, there were three of the 25 patient participants who only completed one story and then asked to withdraw from the study before being interviewed or completing their second story. It is possible that they might have had a more negative experience than the other participants, but their reasons for withdrawing were not given and the only available data is their first recorded story. In answer to the question about first reactions to the storymaking process, these three participants said:

Well to be quite honest I feel a bit silly [laughs] but, I didn't start off to do this story so I've probably learned something along the way. (Tape 128)

[long pause] It's quite hard to come up with ideas, talking about it. (Tape 133)

It would be nice if it came true, wouldn't it (Tape 139)

There is nothing obvious here to distinguish these responses from those of participants who agreed to continue. The first two said they felt silly, or the process was difficult. It may be inferred that the third produced some kind of wish-fulfilment story that perhaps seemed impossibly optimistic, which might have been a negative experience. On the other hand the first participant felt they had learned something and conformed closely to the pattern of a successful story arc. The worst that can be said is that of the 25 patients recruited, 3 withdrew after completing their first story for reasons which can only be guessed at, but may have included dissatisfaction with the storymaking process.

Some of the clinicians who were taught the 6PSM by the researcher did experience visibly strong emotions while they were undertaking the process. On at least two occasions clinician participants became tearful while telling their story in the training session. However this did not seem to be linked with a negative evaluation of the whole 6PSM; on the contrary, these clinicians were likely to evaluate the 6PSM as highly relevant and as a powerful tool:

It was difficult, a real shock that it affected me in that way. Just how powerful it is. I know I said to you in the beginning, oh it's a real boring story and I honestly had no idea that this was anything about me and I was struggling, I thought, to find a story and I thought this was a fairy type story. And from starting to talk about it, it seems quite pertinent.

[Researcher] And are there bits of you in there somewhere?

Yes, but I don't want to start talking about it now. (Laughs.)

(Clinician C7)

This was also true of patient participants who, if they made negative comments, tended to do so about discrete elements of the 6PSM rather than the experience as a whole. Several participants' responses described feeling silly or childish, and this was often in response to being asked to draw:

What's bad about it is that I'm a crap drawer. My drawing looks like a three or four-year old. I can't draw. That's what's bad about it. (Q003)

The drawing wasn't very good. (crocodile)

I found it quite difficult, I found that I thought I was making up a bit of a silly story really. (C022)

In summary on the question of acceptability, more participants described positive experiences than negative, and even those negative experiences that were reported tended to be restricted elements in an experience that was overall reported as positive. There is the possibility that negative feedback was not given because of a desire to please the interviewer, but even if negative opinions were held but not expressed, no participant became so distressed by the process that it was halted or abandoned. The three patients who withdrew from the study after their first story may have had a more negative experience, or they may have withdrawn for other reasons.

15.2 A THEORY FOR THE OPERATION OF THE 6PSM

Following on from the previous section, there may be data of importance in the negative reactions that were expressed, in particular the feelings of childishness and silliness. Perhaps participants' instinctive reaction is to do with two things. Many people's memory of being asked to draw things is probably bound up with last being faced with this task at school. There may be associations of being judged or marked, of potentially being laughed at or humiliated, or of having one's efforts compared unfavourably to others.

Drawing and storytelling are activities that, for most people, are confined to childhood and that decline or disappear as adulthood is reached. For some people a return to childlike ways of communicating may be liberating and enjoyable; a return to the play space described by Winnicott (1971), with its connotations of the infant (storyteller) being securely held by the parent figure (listener). At the same time the return to childhood may be a return to embarrassment, powerlessness and shame; these may be difficult and unpleasant feelings to return to, but all the more necessary for that. The experience of the 6PSM seems to provoke such feelings yet in a manageable way.

A parallel possibility is that participants have fantasies about the communication of unconscious material and the potential of the 6PSM to make them reveal more than they intend to. Although, as discussed in the previous chapter, nobody used the word 'unconscious' in their discussion of the 6PSM, several participants made reference to potentially revealing more than they intended:

...you can see yourself trapping yourself into themes from you own daily life (Clinician C13)

...it's a bit scary because you've got to be very careful about what's safe and what isn't, because it always ends up being a lot bigger than the pictures that you first drew (Clinician C17)

Most of it feels a bit exposing. (Patient B8, borderline diagnosis)

It has already taken my mind off what I came here thinking, and I'm already thinking what that means, and quite worried about what it does mean (laughs.) (Patient M4, no BPD diagnosis)

A too self-revealing process (Clinician C24)

However the revelations that were feared appeared to be things already known to the participant that they were fearful of revealing to others; almost no participants described revelations that were new to them. Only one participant described the process as giving them a new perspective on themselves to consider:

...I wouldn't have said that I was so distrustful, but it makes me question whether I am or not. (Clinician C21)

Everybody else seemed to regard the 6PSM as a process for communication to others, rather than for revelation to oneself. The worry was that the extent of the

communication might not be as controllable in the 6PSM as in ordinary conversation.

15.2.1 GROUNDED THEORY AND THE CENTRAL CATEGORY

The central category emerging from the Grounded Theory analysis of transcripts was that of Progressive development of story. Emerging from this came the two different profiles of those participants with strong/positive reactions and those with weak/negative reactions, suggesting that the former group had a dynamic experience of the 6PSM process, while the latter group's experience was static. This may parallel the two findings described by Rennie (1994) who undertook another Grounded Theory analysis of the experience of clients telling stories in psychotherapy.

Rennie observed that there were two major functions of storytelling described by the clients; (1) as a means of distancing themselves from their emotional disturbance and (2) as a therapeutic experience. The first function served to delay or prevent active therapeutic work on the emotional disturbance. This is perhaps rather like the participants in the present study who had a static storytelling experience, with a main character who remained distant and where the level of emotional intensity was constant and low (Rennie, 1994: p237-239).

Rennie described the second function in three ways. First, participants reported emotional relief and catharsis. Secondly storytelling brought participants into useful contact with emotionally disturbing material, images, and memories. Third, storytelling allowed participants to process the private thoughts, feelings

and images associated with the story (Rennie, 1994: p239-240). All three elements of Rennie's second function seem to occur in the present study, although the word *catharsis* itself is not in fact used by any participant. Perhaps the dynamic story arc experienced by many participants is the key. Increasing identification with the main character leads to an increased level of emotional arousal, with the nature of the emotion itself changing from anxiety (about what is to come) to upset/emotional (as the links with the teller's own story become clear). As the story is resolved and finally discussed, the identification with the main character leads to an end emotional state of calm and satisfaction. This is very close to the original Aristotelian concept of dramatic catharsis. This is said to arise through purging the audience's emotions via the pity and fear evoked on behalf of the protagonist of the drama (Aristotle, 350BC). The dynamic story arc of the 6PSM facilitates this catharsis in some participants, while the static line of the uninvolved storytellers prevents the identification, the arousal and the catharsis that follows the return from the aroused state for others.

It could even be said that the 6PSM works for both groups of participants; those who are in Rennie's second group can use the storytelling productively to bring them closer to their own material and to process that material with satisfying results. Those in Rennie's first group can use the storytelling defensively to maintain their distance from their own material. This is very strongly reminiscent of the description of aesthetic distancing described by Jones (1993), wherein the metaphorical nature of the dramatic medium allows the client to move at will along the dimension of personal involvement with the material. Those clients who fear they may be overwhelmed by their material can keep it distant and

minimise their personal identification; those who want to become more involved may be helped to do so by the metaphor.

15.2.2 EXISTING MODELS FOR UNDERSTANDING THE ACTION OF THE 6PSM

Stiles, Honos-Webb & Lani (1999: p1217-1223) enumerate five possible functions of narrative in psychotherapy:

1. A distraction or defensive manoeuvre to avoid anxiety
2. An aid to the emergence of warded-off material
3. A strategy to suppress unwanted thoughts
4. A representative of an unwanted voice
5. A way of constructing an understanding

Their study (like almost all studies of narrative in psychotherapy) focuses on personal, autobiographical narrative in therapy, not fictional narrative such as that produced by the 6PSM or other projective approaches. Nevertheless these studies are the closest analogues of fictional storymaking in the literature.

The first and third of the above functions may have been employed by some of the participants in this study. Those participants who found themselves uninvolved and with a flat storyline might have been avoiding anxiety by using the story as a distraction. Alternatively, this group might have been participants for whom this story, at this time, performed no function at all; they were telling the story for the researcher's benefit, at the researcher's request rather than for any reasons of their own.

The second function, aiding the emergence of warded-off material, was described by almost nobody. There was one participant who was reminded by her story of an incident from her own childhood where she was attacked. She described the incident to the interviewer thus:

I feel as though it evokes stuff into your head, it evokes other things. I remember going out in fishnets and black, and I have never told you this and getting in a man's car and I don't even know to this day he was old enough to be my dad and he attacked me, he wouldn't let me out of the car and he drove all the way to the bus station, he said his daughter is my age and he begged me to forgive him and I just humbled myself with him. And I've never told anybody about that, ever. (Patient M11, no BPD diagnosis)

However in this case the patient makes it clear that she herself has always remembered this memory; it is hardly the return of repressed material. Perhaps the second function, the emergence of warded-off material, is performed here if this material is seen as having been warded off from the clinician, rather than the patient. But even this was the sole example of such a revelation.

There were a number of stories that might have been fulfilling the fourth function, representing an unwanted voice. As an example, one patient participant told a story of a teddy bear trapped in a room, trying to get out but hurting himself in the process. This story of a vulnerable, weak, needy main character came from a patient with a diagnosis of Antisocial Personality Disorder who presented in interview as very controlled, confident and assertive.

This story may have been his attempt to communicate or identify an unwanted voice, but to have confirmed this would have required a much more flexible and exploratory response style from the interviewer. Instead, to give participants the same stimulus, all interviewers used a fixed script of general questions and were asked not to embark on the kind of interpretive discussion that would be necessary to find stories embodying this fourth function.

The fifth function, constructing an understanding, seemed to be present more frequently. One patient said:

...I don't think it was a waste of time because maybe it's me try to tell myself that I'm wasting my life and there are things I could do about it.

(Patient B6, borderline diagnosis)

While one of the clinician participants said:

... the last bit surprised me. Because I am a kind of happy ever after person really but there is something about survival and I think for me that is probably the most significant part of the story, it is about surviving when things are difficult. (Clinician C6)

Many participants made similar comments, probably prompted by the question in the interview script that asked:

And what if this story was like one of those parables or fables, that as well as being a story has some kind of teaching in it. What lesson, moral or advice does the story have for you?

This question does seem to be explicitly inviting participants to construct an understanding of themselves via the medium of the story. Perhaps because of this explicit question, the fifth of the functions outlined by Stiles et al. (1999) seems to be more in evidence in these transcripts than any of the other four. It may be that the 6PSM is less a tool for the unearthing of previously unknown material, and more a tool for the reorganisation of existing knowledge about the teller.

15.2.3 CONSTRUCTING AN UNDERSTANDING - A COGNITIVE MODEL

So perhaps one end-point of the 6PSM is the construction of an understanding.

Among the building blocks for this construction are:

- Instructions heard by the participant transferred into mental images
- Mental images transferred into drawn images
- Drawn images subsequently acting as prompts for spoken responses
- Reflection upon the spoken responses leading to the construction of understanding.

This repeated transformation of material from one form to another was very strongly reminiscent of the process described by Teasdale & Barnard (1995) in their description of the Interacting Cognitive Subsystems (ICS) framework. This model, drawing on both experimental and clinical data, proposes that there is a number of interlocking information processes at work in processing any task. Each process receives inputs, makes recordings and transformations and produces outputs, which themselves may form inputs for the next stage of the process.

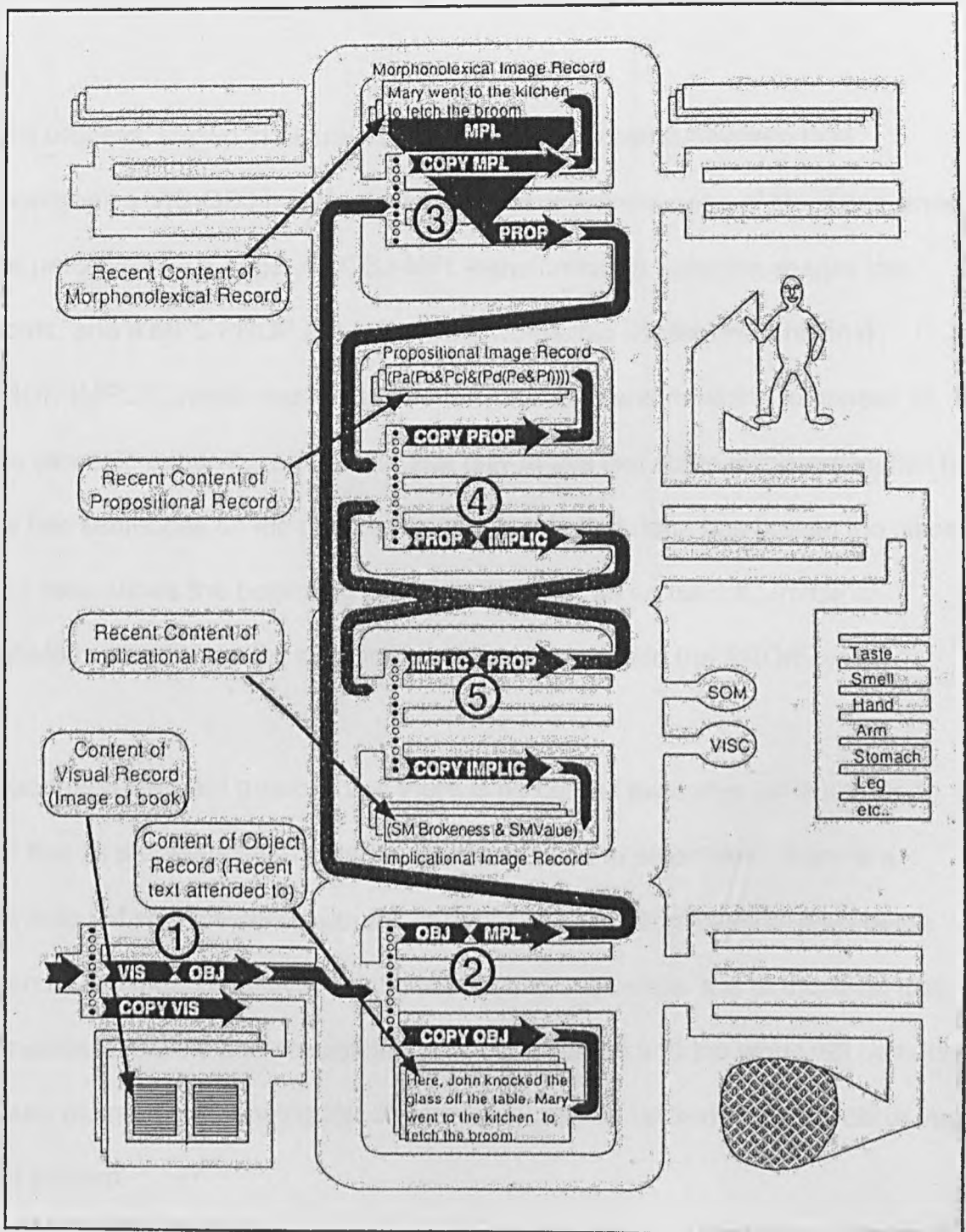
The nine components of the ICS, with their functions are as follows:

Table 15.1: Components of the Interacting Cognitive Subsystems model

Subsystem	Nature of information recorded and processed
Sensory & proprioceptive subsystems:	
Acoustic (AC)	Pure sounds from acoustic nerves - what we 'hear in the world'
Visual (VIS)	Visual stimuli from optic nerves - what we 'see in the world'
Body state (BS)	Stimuli from other sensory nerves, proprioception etc - tastes, smells, pain etc
Systems of intermediate structural description	
Morphonolexical (MPL)	Sounds interpreted as words - what we 'hear in the head'
Object (OBJ)	Visual stimuli understood as objects in relationship - perceived as 'visual imagery'
Meaning systems	
Propositional (PROP)	Semantic descriptions, experienced as 'knowing that...'
Implicational (IMPLIC)	High level codes derived by inference, familiarity, causal relatedness
Effector subsystems	
Articulatory (ART)	Decisions and plans for motor action leading to speech - subvocal speech output
Limb (LIM)	Decisions and plans for motor action leading to body movement, 'mental' movement

Each subsystem is linked to others by a system of inputs and outputs that is complicated, but elegantly described by Teasdale & Barnard (1995: p59). Their diagram describing one sequence of information processing is overleaf as Figure 15.1.

Figure 15.1: The "central engine" in text comprehension (Teasdale & Barnard, 1995: p77)



Teasdale & Barnard give as an example the activity of reading a piece of text such as

John knocked the glass off the table. Mary went to the kitchen to fetch the broom.

This process, shown in Figure 15.1, starts in the visual subsystem then undergoes a VIS-OBJ transformation so that ordered shapes of black and white are perceived on a page. An OBJ-MPL transformation turns the shapes into words, and a MPL-PROP assembles the words into sentences. The final PROP-IMPLIC transformation allows interpretation and meaning to appear. In this example the authors point out that only at this last stage is the implication of the two sentences on the page realised; namely that John has broken the glass. This then allows the beginning of the next phase, as an IMPLIC-PROP transformation brings the proposition of brokenness into the PROP system.

Teasdale & Barnard propose that there is no central executive for this system, but that all elements interrelate as a network. On the other hand, there is a hierarchy of sophistication with the information at the Implicational level being seen as the most high-level, sophisticated and schematic. It is at this level that generalised beliefs and predictions about self, others and the world will be held, based on information being transformed and passed up from lower levels of the ICS system.

These transformation processes are characterised in four ways:

1. They 'learn' the appropriate recodings from one code to another on the basis of repeated experiences of co-occurrences between input patterns and output patterns
2. They can be viewed as embodying 'procedural knowledge' in that they 'know' what output pattern to produce given a particular input pattern
3. They are sensitive to the total pattern of input code so that a few discrepant elements in the total pattern can have profound effects on the output; but
4. They can operate on partial or degraded input patterns so long as these do not contain discrepant elements.

(Teasdale & Barnard, 1995: p55-56)

This model suggests that unambiguous or familiar inputs will be processed quickly and lead to a predictable output, while strange or ambiguous inputs may lead to no output, or to a 'best fit' output on the basis of whatever is in the relevant information store. One explanation of the action of the 6PSM (or any other projective technique) is suggested by the fourth point above. Inputs such as a TAT card or Rorschach blot can be seen as partial or degraded; the output therefrom is likely to be predicted by the 'procedural knowledge' described above (point two), which itself will be dependent on the repeated inputs experienced up to that time (point one).

This may explain the repeated descriptions from participants in this study of the experience of things becoming more relevant and focussed as the storymaking exercise went on. At the beginning they were presented with the stimulus of a blank sheet of paper (VIS) and an instruction to draw a main character

(AC→MPL→PROP). These extremely partial inputs led to the output of a drawing (PROP→VIS→LIM) along with the beginning of an as yet unspoken story (PROP→MPL→VOC). At this stage there is no real role for the Implicational system; these are images which have little meaning and no inferences can be drawn from them. Perhaps this is why many people reported that this stage felt trivial:

I found that I thought I was making up a bit of a silly story really.

(Clinician C18)

However, once the drawings are in place and the participant is asked to start to tell the story the inputs are less vague and more complete; rather than a blank page, there is to begin with a set of images which can activate the VIS→OBJ→PROP→IMPLIC systems. Interestingly, this bypasses any verbal levels of information processing and may represent for most people a less-used (and perhaps less defended) route. As the story is told and then questions are asked about it, it seems likely that the Implicational subsystem is more and more activated; some of the later questions specifically ask the teller about what the story means. Such questions demand answers that can only be answered either by a reference to the Implicational subsystem or by an assertion, made by some participants, that it means nothing. Arguably even these participants had to access the Implicational subsystem to establish that it contained no data.

The ICS approach has echoes of the Gestalt psychology of the late 19th and early 20th century; one of the pioneers of which was Maria Rickers-Ovsiankina. She subsequently became President of the Rorschach Institute in the USA and noted the debt that the Rorschach system owed to Gestalt Psychology (Rickers-Ovsiankina, 1992, originally published 1943). In particular she pointed to the

work of Kurt Lewin (1936). One of the best-known elements of Gestalt Psychology was the attention paid to the active organising processes involved in perception; for example the idea of figure completion. Here an incomplete, unfamiliar stimulus is perceived as a complete, familiar figure. In the example below, a central white triangle is usually perceived even though the marks on the paper actually consist of six separate black elements.

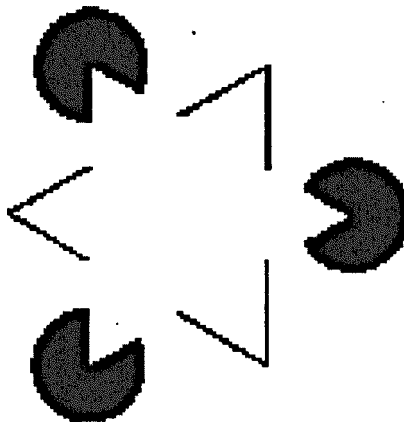


Figure 15.2: Gestalt figure completion

The idea is that the cognitive process makes the best sense it can of the unusual inputs, organising them into the simplest and most familiar explanation. So the figure above is usually seen as a white triangle overlying a black-outlined triangle and three circles.

Similarly, when presented with the vague and unfamiliar stimulus of the instructions of a projective technique such as the 6PSM, it may be that simple and familiar mental images are elicited. These then form the ground against which further images more easily appear. The concept of figure and ground is another familiar one from Gestalt Psychology; most frequently seen in the

familiar vase/two faces illusion described by the Danish psychologist Edgar Rubin (1915) and reproduced below.

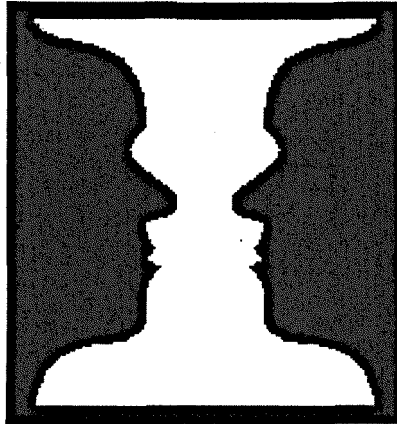


Figure 15.3: Figure-Ground in the Rubin illusion

In projective techniques such as the 6PSM, the first responses given by the participant may become the ground against which subsequent responses figure. Thus the story becomes clearer and firmer by this iterative process of a prompting question bringing something into the foreground, figural position after which subsequent questions relegate it to the background, allowing the emergence of a new figure.

16.0 FINAL DISCUSSION

16.1 INTRODUCTION AND REVIEW OF RESULTS

This chapter will begin with an overall review of the results of the research, followed by discussion of the limitations and shortcomings of this study. A review of the implications for clinical practice will follow and the thesis will end with indications for future research.

The results of the research will be reviewed with reference to the research questions outlined at the start of the study and described in Chapter 5, namely:

- In the context of people with a personality disorder, can a practical rating technique for six-part stories be devised?
- Given this system, can two raters independently rate stories reliably?
- Are the ratings of stories from the same authors reasonably stable over time, with the possible implication that they are tapping relatively stable characteristics rather than more transient states?
- Do the ratings given by raters to others' stories bear any relation to the personal characteristics of the rater themselves?
- Are the results from the rating method/s devised for the 6PSM associated with any other data about the stories' authors - for example personality disorder diagnosis, general mental health or gender.
- How acceptable is the 6PSM to patients as a method for general clinical use?
- To what extent does the 6PSM produce material that has face validity for the authors themselves?

- Do the subjective accounts of authors of a six-part story permit the development of a theory about the mode of action of the 6PSM?

16.1.1 CAN A PRACTICAL RATING TECHNIQUE FOR SIX-PART STORIES BE DEVISED?

The rating form for the pilot study, described in Chapter 6, would not have been usable as a routine clinical or research instrument, nor was it designed as such. The form was a lengthy (8 page) document that took respondents between 45 minutes and 3 hours to complete for a single story. The fact that tape recordings of 6-part stories were provided may have added to the completeness of the information available to respondents, but it certainly also added to the time taken to review and rate a story.

Of the several methods of rating in the pilot form, the most intriguing to the author was the least successful; this was the attempt to derive a set of reciprocal roles from a repertory grid style analysis of actants in the story. In 6PSM training sessions run before the study started, participants had always been interested by the potential of the technique. This was particularly true of Cognitive Analytic Therapy trainees, for whom reciprocal roles are a stock-in-trade, but even these trainees found the completion of the grid confusing and over-elaborate. Participants in the pilot research were not able to agree among themselves as to the descriptions to be given to relationships between actants.

In clinical practice this method has been used by the author in collaboration with the patient telling the story; the grid (a simplified 3x3 version) is constructed by

the clinician but the identification of actants and their relationships is carried out by the patient. This may be a more clinically fruitful procedure, and certainly leads to a high level of face validity from the patient's point of view, but is not susceptible to the same kind of inter-rater reliability testing precisely because the actants and their relationships will have a highly personal meaning for every storyteller. This procedure treats the 6PSM as a method of gathering idiographic clinical material that is unique to the patient, rather than material that would allow comparison with any other patient or group.

The method of rating the 6-part story with the most promise for comparing patients and groups is the rating of statements about stories. This was quick and easy for clinicians to fill in, and now that a number of unreliable items have been removed the scale for potential clinical use is even shorter. The group of statements was based partly on predictions from theory or practice, and partly on the actual content of some of the stories collected. It is highly likely that there are other statements about stories that would also reliably allow raters to distinguish between groups, if further research on a new set of stories were undertaken.

16.1.2 CAN TWO RATERS INDEPENDENTLY RATE STORIES RELIABLY?

Fourteen statements describing stories were found to have acceptable inter-rater reliability. From these, two sets of statements were identified on which raters could agree reliably. One group of statements related to the degree of negativity and pessimism expressed in the story and the second to the degree of violence present in the stories. As might be expected, stories from

participants with a cluster B personality disorder were more negative. There were fewer helpful others in stories from this group, which is perhaps not surprising except that the stories might have been used as fantasies of wish-fulfilment containing idealised, rescuing others: this did not seem to have been the case.

Agreement on the Core Conflictual Relations Themes (CCRT) was also good, once the eight CCRT themes were collapsed into two simple positive-negative categories. Very few stories were rated as having a negative wish for the main character, so this was not a helpful category. The other two themes – response of others to wish and response of main character to others – could be agreed upon reliably by two raters.

The single rating of global mental health, as inferred from the story, could also be made reliably by two raters. There were sufficient ratings of this to enable the performance of more and less experienced raters to be compared. More experienced raters (those who had experienced 5 or more stories in a clinical or training context) were more likely to agree with one another than less experienced raters, but even the latter achieved a significant level of agreement.

This research question could be answered positively; there were indeed several measures on which two or more raters could reach significant levels of agreement.

16.1.3 ARE RATINGS OF DIFFERENT STORIES FROM THE SAME PARTICIPANTS STABLE OVER TIME?

Not all of those ratings with inter-rater reliability could be shown to have test-retest reliability. Of the two sets of statements identified in stories, those relating to the amount of violence in the story were not stable over time. Thus although raters could agree on these statements at both times, these factors fluctuated over the one-month interval between stories and are unlikely to be related to an underlying personality disorder which could be assumed to be more stable. These statements may relate to more transient conditions such as mood, life events or subjective well-being.

The statements relating to the degree of negativity however appeared to be stable over time. When pooled into a single scale these five statements had a moderate test-retest correlation (Intraclass correlation = .49, $p < .05$, $df = 16$) which, together with their strong inter-rater agreement (Intraclass correlation = .85, $p < .001$, $df = 54$) made them promising candidates for a scale that might relate to personality disorder.

Of the two CCRT themes with acceptable inter-rater reliability only that of the reaction of the main character to others was stable over time. This test-retest reliability was moderate ($Kappa = .69$, $p < .05$, $n = 17$) as was the inter-rater reliability ($Kappa = .58$, $p < .001$, $n = 34$). This level of reliability is somewhat less than that for the negativity scale derived from story statements, which may be a function of the fact that the CCRT ratings were dichotomous (positive-negative) while the negativity scale had a range of 21 possible values. This

range may make the latter scale a more sensitive and useful measure than the CCRT rating.

The global rating of mental health inferred from the stories was also reasonably stable over time, suggesting that it is tapping something more than a transient phenomenon. As with the CCRT, the reliability was only moderate for both test-retest ($r = .64, p < .01, n = 17$) and inter-rater ($r = .46, p < .001, n = 56$) reliability. Coupled with the fact that it is not immediately clear exactly what this rating is tapping, the negativity scale is probably to be preferred to this global rating.

In addition to the scales derived from participant ratings, there was also the data derived from the computerised text analysis. This of course did not require any measure of inter-rater reliability as it was conducted by routines which could be expected to arrive at the same answer every time! Of the three dictionaries used, the scale derived from the Linguistic Inquiry and Word Count (LIWC) had the best test-retest reliability, but this was only moderate ($r = .62, p < .01, n = 18$). The Regressive Imagery Dictionary and General Inquirer had somewhat lower levels of reliability ($r = .43$ and $.44$ respectively.)

In addition to the measures derived from the text analysis, the simple number of words in the story transcript was also very stable over time ($r = .90, p < .001, n = 18$).

It was therefore possible to answer this research question positively; there were several rating methods that were stable over a one-month test-retest period, all

of which had previously been shown to have acceptable inter-rater reliability (where necessary).

16.1.4 DO THE RATINGS GIVEN BY RATERS TO OTHERS' STORIES BEAR ANY RELATION TO THE PERSONAL CHARACTERISTICS OF THE RATER THEMSELVES?

This question related to the degree to which rater bias might be a factor. For example, would raters whose own stories were rated by others as pessimistic be more (or even less) likely to rate other people's stories as pessimistic?

Sadly, the numbers of raters who had also produced 6-part stories were too few for this question to be answered. Thirty clinicians provided stories, and a similar number provided ratings, but there were only five clinicians who provided both. With this number it was not possible even to detect a trend.

Regrettably therefore this interesting question could not begin to be addressed by the results from this study. The only rater characteristic about which there was any information was the degree of experience with the 6PSM, as discussed in section 16.1.2. More experienced raters are more likely to agree on ratings of 6-part stories, but it is not possible to say whether this is because raters become less biased with experience, or whether with more experience their biases tend to converge.

16.1.5 ARE THE RESULTS FROM THE 6PSM VALIDATED BY ANY OF THE CONCURRENT DATA GATHERED ABOUT PARTICIPANTS?

The mean negativity score derived from statements about the stories was significantly different for stories from those with a diagnosis of a cluster B personality disorder and those without – when clinicians were factored into the calculation as probably *not* having such a diagnosis. There appeared to be no difference in the mean negativity scores of stories from patients without BPD and clinicians. This suggests that it is the presence or absence of BPD, rather than the presence or absence of significant mental health problems that is being picked up by the negativity score.

However, it is not possible to say with certainty whether variations in the negativity score does are more closely linked to Borderline PD pathology than to depressive illness. Further studies sampling populations with and without cluster B disorders and depressive illness would be needed to investigate this further.

There were considerably more stories from women than men with a cluster diagnosis, an imbalance not present in other groups. This did not affect the negativity score however, which was independent of gender and much more closely linked to cluster B diagnosis.

In addition to the negativity score, short (2-3 item) scales were identifiable that seemed to distinguish between participants with diagnoses of Schizoid PD and Avoidant PD and others.

The CCRT rating of the main character's response to others as positive or negative was very different for stories from those with and without a BPD diagnosis. Only 2 of the 12 stories from authors with a BPD diagnosis had a main character with a positive response to others; this is an interesting feature which had not been predicted. From the qualitative data discussed in chapters 14 and 15 it is clear that participants generally see the main character as an expression of themselves, and it might have been expected that the main character would be an idealising, help-seeking, dependent figure in stories from those with a BPD diagnosis. Instead it seems that the stories may be illustrating the more denigrating, attacking pole of the borderline experience.

The rater inference of global mental health was, like the negativity scale, moderately correlated with the total number of SCID-II criteria met ($r = -.63, p < .001, n = 38$). The ratings given to stories from patients without a BPD diagnosis and those from clinicians were not significantly different, while stories from patients with BPD were rated with significantly poorer inferred mental health.

In the computerised text analysis, the LIWC-B score derived from the Linguistic Inquiry and Word Count was the most effective in distinguishing between stories from participants with a BPD diagnosis and others. At the optimum cut-off point, this score had a sensitivity of 68% and a specificity of 91%, correctly identifying the author's BPD diagnosis in 32 out of 40 stories. Among the concurrent measures the LIWC-B scale was correlated most strongly with the CORE Risk subscale ($r = .59, p < .001, n = 38$).

When comparing the scales derived from participant raters and computerised text analysis, the former (negativity scale) seemed to be more closely linked with the general level of personality disturbance as measured by the SCID-II, while the latter (LIWC-B scale) seemed to be more closely linked with the CORE-Risk subscale. The two scales themselves were only weakly correlated ($r = .28, p < .05, n = 64$).

These two scales have good test-retest and inter-rater reliability, they are internally consistent, have face and concurrent validity and distinguish well between stories from different participant groups. They each seem to be tapping somewhat different features of the stories and their authors, and as such could usefully be used in tandem.

16.1.6 IS THE 6PSM ACCEPTABLE TO PATIENTS AS A METHOD FOR GENERAL CLINICAL USE?

Most participants who expressed an opinion were positive about the usefulness and relevance of the 6PSM. Many participants experienced anxiety about the process, described feeling silly or becoming emotional, but this did not necessarily stop them from approving of the overall process. Three of the 25 patient participants withdrew from the study after their first story session, but there is nothing in the transcript of their first session to indicate they found the first session more difficult or less acceptable than other participants. A small number of participants expressed very positive opinions of the 6PSM, saying

that it very quickly and accurately allowed them to communicate parts of themselves in a new way.

16.1.7 DOES THE 6PSM PRODUCE MATERIAL THAT HAS FACE VALIDITY FOR THE AUTHORS THEMSELVES?

The greatest number of comments in the post-story discussions related to the relevance of the story to the teller's own life, and the majority (by about 6 to 1) were positive. Only two of the 49 participants who made positive comments said that the story had no relevance to their own situation. Many participants commented that the story began as irrelevant or trivial, but became increasingly personal and meaningful as the process continued. Participants tended to see the 6PSM as a tool for communication to others, rather than self-discovery.

16.1.8 DO THE PARTICIPANTS' ACCOUNTS PERMIT THE DEVELOPMENT OF A THEORY ABOUT HOW THE 6PSM WORKS?

A theory has been developed about the action of the 6PSM, based on participants' observations that the main character of the story often begins as distant and universal, before becoming increasingly identified with the story's author. This is paralleled by a shift in the emotions experienced, with the structure of the story arc encouraging a deepening of emotional intensity in the midpoint of the story, with a return to a calmer, more reflective place at the end.

The inverse relationship of the aesthetic distance and emotional intensity was experienced both by participants who experienced the process as highly personally relevant and those who did not. The difference is that the former experience a process where aesthetic distance and emotional intensity are constantly changing, while the latter experience no change throughout the process.

The 6PSM process starts with a blank canvas with infinite possibilities and no particular personal relevance. The story building process puts down successive layers of imagery which, because they have been furnished by the teller, usually become more and more personally relevant. This iterative process, translating verbal instructions into drawn images and then words, is well described by and consistent with the mechanisms of the Interactive Cognitive Subsystems approach (Teasdale & Barnard, 1995).

16.2 RELATION TO PREVIOUSLY PUBLISHED RESULTS

There are no published studies of the 6PSM with which to compare these results. However there are studies of other projective approaches with patients with depression and (less frequently) personality disorders. The results from this present study are consistent with previous findings. Westen, Ludolph, Lerner, Ruffins, & Wiss (1990) compared the Thematic Apperception Test responses of adolescents with a diagnosis of BPD and other adolescents. They found that "Borderline adolescents have a malevolent object world [and] a relative incapacity to invest in others in a non-need-gratifying way" (p.355), showing more negative, malevolent relationships and selfish, aggressive impulses than

the former group. The findings were repeated in a parallel study of adults with and without BPD (Westen, D, Lohr, N. E. *et al.*, 1990). This is consistent with the present study, which found a consistently negative cast to the main character's response to others.

The negativity factor identified in this study raised the question of the degree of depressive illness among the sample studied. Other studies of projective tests have found that depression is linked with negative outcomes in projected stories as well as hostile feelings towards others (Holmstrom, Karp, & Silber, 1994). However their study, while taking a measure of depression, did not assess Axis II pathology, so it is not possible to say which was having a greater influence on negativity and hostility in stories.

Huprich (2001) found that a profoundly negative, pessimistic outlook was more common among subjects with dysthymia and depressive personality disorder than those with a dependent personality disorder. In the present study, only one of the patient participants had a diagnosis dependent personality disorder. On the other hand, eight of the 11 participants with a diagnosis of BPD also had a diagnosis of Depressive PD, suggesting that this group may be similar to that in the Huprich study.

One other study (Ackerman, Clemence, Weatherill, & Hilsenroth, 1999) also compared the TAT responses of participants with BPD and other personality disorders. They found that stories from participants with BPD had significantly poorer (malevolent, negative or abusive) relationships in their stories; a finding which is consistent with the present study.

16.3 LIMITATIONS OF PRESENT STUDY

This study has attempted to use two different methodologies that are not usually seen together; quantitative approaches necessary to establish psychometric properties are not often employed in tandem with a Grounded Theory analysis. During the period of the research, many people commented to the author with surprise that he was using an essentially quantitative method to study a projective tool producing such rich qualitative text data. The study could easily be criticised by methodological purists of either school as deviating from the usual field of enquiry. Even a charitable view may be that such a hybrid project might have all the weaknesses of both parents with the strengths of neither.

In this respect the pre-eminence given to the quantitative data collection and analysis means the gathering of qualitative data suffered. Interview data from clinicians could be gathered directly by the researcher from clinicians in training them to use the 6PSM, but this was not the case for patients. In order to maximise the uniformity of prompts given to patients, clinicians used scripted questions to elicit the 6-part stories from patients, and this extended to the questions about the process itself. It was not possible therefore for the answers provided by patients to be explored *in vivo*, or for unexpected and interesting lines of enquiry to be pursued. It would have been possible, because data were collected over several months, to rework the end of the script to explore the reactions of later patients in the light of those reported by earlier patients. However this opportunity was not taken.

On the quantitative side, the numbers involved were much smaller than had been hoped. Sufficient clinicians were recruited, but they themselves were not able to recruit nearly as many CMHT patients as had been envisaged. Instead of 30 clinicians, 60 patients and 150 stories there were 24 clinicians, 25 patients and 65 stories. Had the target number been reached it is possible that further levels of significance and finer distinctions (particularly perhaps in the computerised textual analysis) might have been drawn.

For example, there were no statistically significant differences found between stories from male and female participants. This may be a Type II error arising because of the small numbers. No examination of other subgroups (for example by age) was attempted because of the low numbers involved in any such groups.

The concurrent data gathered did not include CORE, IIP or SCID-II data from the clinicians, which would have allowed a wider range of data to be gathered. In addition, although the clinicians were never formally established as a control group, they were compared to the patient groups as if they formed some kind of non-patient controls. However they were a very homogeneous group as regards age, education, gender and social status. It would have been better to recruit a fourth group of people from the general (non-mental health service using) population to act as controls, with all the concurrent data gathered on them.

The concurrent data did not include a measure of Axis I disorders, so it is not possible to tell how far this may have influenced the results. For example, if

most of the patients with a BPD diagnosis were depressed, but most of the mainstream CMHT patients were anxious, this would seriously confound the results.

The recruitment of patient participants was done on the basis of convenience and availability, rather than through any purposive sampling strategy. Although some randomisation was undertaken when selecting from a list of potential candidates, this list was potentially highly selected by the clinician. There is no guarantee that the mainstream CMHT patients recruited to the study are in fact representative of the distribution of patients on a CMHT caseload.

An associated weakness arises from the location of the research, which took place in an area with very little ethnic diversity. As a result all the participants (both patients and clinicians) were white UK nationals. The sample accurately represents the local population in this respect, and it could be seen as a strength that the sample were relatively homogeneous in this regard. However, in a field susceptible to cultural influence as the telling of stories, it is a shame that there was so little diversity in the participants' backgrounds.

The scales derived from the rater and textual analysis suffer because they have been tested on the same group from which they were derived. This will capitalise on any random variations in the data and make it more likely that levels of significance may be artificially high. It would have been better to derive the scales from one set of stories and patients, then test their sensitivity and specificity on a separate set of stories. Unfortunately the small numbers involved in this study made a split-half procedure of this sort impossible.

The rater inference of mental health is a vague and general concept. Although it appears to have inter-rater and test-retest reliability, it is difficult to understand quite what it is this rating is measuring.

The 6PSM as applied in this research is not the form in which it is used clinically. Most significantly, the script which was used to standardise the prompts given to participants is not recommended in clinical use precisely because the clinician is intended to ask questions that are tailored to and arise from the specific story being told. This means that, strictly speaking, any use of these results in a clinical setting would have to be treated with caution unless the script and procedure from this study were used.

The text analysis procedures used in this study require the taping and transcription of stories which is necessarily a time consuming business. Although such transcriptions are clinically valuable, it is unlikely that many clinical settings could afford the time to do this routinely. Moreover, the text analysis software necessary to analyse the transcripts is extremely unlikely to be available in a clinical setting.

16.4 IMPLICATIONS FOR CLINICAL PRACTICE

As described in Section 4.2.2.2 the 6PSM has been widely used by dramatherapists and others in the UK, USA and Israel with no evidence base for its use. Clinicians can now use the 6PSM in the knowledge that its validity and reliability has been tested in at least one area.

16.4.1 TRAINING IMPLICATIONS

The clinicians who recorded the 6-part stories received less than three hours of training, involving creating and telling their own story and then listening to and questioning a colleague on theirs. A few of the clinicians had previously received longer training from the author, but for the majority of clinician participants this was their only experience of the 6PSM. Nevertheless they were able to conduct sessions with patients producing 6-part stories, stories which most of the patient participants felt to be relevant and sometimes powerful.

Most of the raters who provided the blind ratings of stories had received a little more training - perhaps one working day - and may have been slightly more experienced in the 6PSM than the clinicians eliciting stories. They did not however receive any training in the rating methods employed; they simply followed the written instructions they had been given.

This suggests that it is practical to train clinicians in using the 6PSM within a single working day, following which they would be able to elicit, discuss and rate a 6-part story according to an established protocol. One element of training that

has not been possible until now would be to give participants one or more sample 6-part stories and invite them to rate these using the methods developed here. The results do suggest that slightly more accurate ratings (in terms of inter-rater agreement) can be provided by clinicians who have been exposed to 5 or more stories, so it is likely that the quality of ratings will improve with experience.

If the data from this study were to be used in any way as normative for clinical purposes then there would need to be close adherence to the script and questions developed and used in this study. There is a tension here between capitalising on the flexible, creative potential of the 6PSM and remaining close to the form which has been validated in this study. Initial training should certainly be followed by supervision (possibly peer supervision) in the 6PSM. It might also be helpful to consider periodic re-training of practitioners who use the method to ensure continued adherence to the protocol.

16.4.2 PRACTICE IMPLICATIONS

As a rule of thumb, the better the mental health and the fewer the personality difficulties of a 6-part story author, the more optimistic and successful their story is likely to be. This bears out the impression gained that trainees, clinicians and others who are not using mental health services tend to produce vivid, coherent stories with a successful main character. What this research has shown is that CMHT patients without a personality disorder also produce similarly successful, optimistic stories, so the 6PSM in this form and with these scoring systems may not be sensitive to the presence or absence of Axis I disorders. The thrust of

this research has been in the context of personality disorder and it is only in this area that the results could be seen as both valid and reliable.

In more general mental health settings, or in settings other than mental health (such as education or personnel settings, where the 6PSM has been used in Israel) the method may be reliable but the validity of any results gained must be less certain. Moreover, this study has concentrated on adults in the 18-65 age range, so no conclusions can be drawn about the stories that might be told by children, adolescents or older people. As mentioned in section 16.2, the research was conducted with participants who were all from a white UK background, so generalising to adults from other cultural groups and traditions may be problematic until more work has been undertaken.

Despite these caveats about generalisation, it has been shown that the 6PSM can be applied in a consistent manner via an agreed protocol to produce valid and reliable results. There is no reason to think that personality disorder is the only clinical area in which it will be useful; this has merely been the first area to be investigated.

However it is not proposed that the 6PSM should be principally used as a diagnostic tool in a similar way to the SCID-II or MCMI. Rather it is suggested that the 6PSM could be used as it currently is by most practitioners - as a tool for generating rich, idiographic data - with the additional use of this rating form as a further way of making use of the story information produced. The 6PSM data should be seen as subordinate to data gathered from the patient, other

informants, the clinical impression, self-report etc, but a useful triangulatory adjunct nevertheless.

16.4.2.1 What do the idiographic data reveal about the borderline situation?

It had been thought that different elements of the borderline experience might be illustrated in people's 6-part stories; for example there might have been fantasies of perfect care, wish-fulfilment stories of success or fusion with an idealised other. This did not seem to be the case; all the stories from patients with a BPD diagnosis seemed to come from the place of abandonment depression (Manfield, 1992). This is normally the place against which the borderline defences are so carefully erected, and it may be that the 6PSM offers a way of exploring that place without the need for the immediate re-erection of those defences. The medium of the story allows a degree of aesthetic distancing which permits the negative, depressing, lonely, abandoned state to be experienced at one remove. This may be a useful phenomenon to be capitalised on according to the modality of the therapist.

For example the creative therapist (drama, art, music or dance movement) would have no difficulty in staying in the metaphor and exploring the images produced by the 6PSM through their preferred medium. The psychodramatist might enact elements of the story - either maintaining their fictional distance, or perhaps by asking the patient to be explicit about which real-life events and personalities map onto those of the story. Any approach that values the kind of images produced in dream work might be able to apply the same procedures to the images produced by the 6PSM.

16.5 IMPLICATIONS FOR FURTHER RESEARCH

16.5.1 CONFIRMATION OF VALIDITY

As previously mentioned, the negativity scale and other measures have been tested out on the same sample from which they were derived. This is far from ideal, and their validity can only be properly tested against new 6-part stories from similar groups to those in this study. A confirmatory study to test this would allow more weight to be placed on the results gained in this study. If carried out in a setting where personality disorders are also assessed, features of the 6PSM associated with different Axis II disorders might be made clearer.

16.5.2 EXTENSION OF NORMATIVE DATA

No data from a representative sample from the general population have been gathered, and it would be useful to have such data on what a 'normal' 6-part story might look like. This might allow the differences in the 6-part stories of people with a wider range of mental health problems to be identified more clearly.

If more stories were gathered along with the matching concurrent data, further subgroup analysis might be possible. In particular, any of the gaps identified elsewhere in this chapter might be filled. For example, in what way do children's stories differ from those of adolescents and adults? Can a developmental sequence be identified similar to that for children's drawings? Given that the

6PSM is widely used in child and adolescent settings this is an important question.

It would also be interesting to apply the 6PSM to different cultural groups; the author has taught the 6PSM to practitioners of various backgrounds, and their impression has been that the story process is a near-universally understood one. It seems it would be possible to use the 6PSM immediately with a wide range of patients without, for example, making any assumptions about psychological-mindedness, western models of health and illness and so on. However this impression needs testing out empirically.

A further extension would be to enquire about the Axis I symptomatology of any future research participants, with a view to seeing, for example, whether the 6-part story of the depressed or anorexic patient has any distinguishing features.

16.5.3 ASSESSMENT OF RATER BIAS

It was unfortunate that no assessment of this was possible in this study. The degree of clinician projection involved in the analysis of a story remains an unknown factor. The fact that some measure of inter-rater reliability was possible in this study suggests that at least some of the themes identified in the stories were not purely projected by the rater, but this is a question that still needs an answer.

16.5.4 ANALYSIS BY AUTHORS

In clinical practice with the 6PSM the understanding of the story and the process of examining it in detail is not carried out by a blind rater but by the clinician and the patient who produced the story. When training clinicians, the author emphasises that the patient, not the clinician, should usually be the first to offer an interpretation of the story material. A future study that examined the way patients rate their own stories would be interesting; do they agree with the ratings given by blind raters or do they differ? And would there be a way of assessing whether the patient or the clinician was making a more accurate rating?

As previously mentioned, the author is now inviting patients outside the research setting to use the CCRT to examine in more detail the relationships within the story. Individual patients have reacted very positively to this, and it would be an interesting method to pursue as well as being more true to the usual clinical use of the 6PSM.

16.5.5 EFFECTS ON ALLIANCE

It has been noted that projective techniques run the risk of bringing the Barnum effect into play. Hopefully this study has demonstrated that the 6PSM does not rely entirely on general, non-specific assessments that could apply to anyone. However it is still possible that the procedure might involve some element of the Barnum effect: this need not be seen as an entirely bad thing. Like the placebo effect, the Barnum effect is usually seen as an interesting but not a useful

phenomenon. However, it might be possible to capitalise on the effect by seeing the 6PSM not only as an assessment and communication tool but also as a tool to improve the therapeutic alliance.

The 6PSM seems to be experienced as novel, enjoyable and powerful by patients and to quickly produce highly relevant material. It may be that such an experience has a positive effect on the therapeutic alliance beyond its usefulness in assessment. This could be assessed by using one of the many tools for assessing therapeutic alliance to track therapeutic relationships, and seeing whether there is any significant change in alliance after a session using the 6PSM.

16.5.6 EFFECTS ON THERAPEUTIC OUTCOME

Better assessments and improved alliance should lead to a better therapeutic outcome; indeed it could be said that whatever its mode of action, if the use of the 6PSM does not in some manner lead to an improved outcome, then it is superfluous. In theory it might be possible to detect the effect of using a 6PSM assessment through a large enough trial, but there will be so many confounding sources of variation this is probably a vain hope.

However the Story Evocation Technique on which the 6PSM is based is widely used in dramatherapy and associated fields as a therapeutic medium, not just an assessment tool. Many books attest to the popularity of storymaking as a therapeutic activity (Crimmens, 1998; Dwivedi, 1997; Gersie, 1997) but empirical tests of its usefulness are absent. In section 16.4.2.2 some

suggestions were made about how therapists from different disciplines might take the 6PSM material and use it as raw material for the therapy itself.

The longer term goal would be to see whether the ancient, universal activity of storytelling can be shown to be efficacious and effective in treating mental distress; hopefully this study has been an early step along that road.

17.0: APPENDICES

APPENDIX 1: PILOT RESPONSE FORM

PLEASE DO NOT READ THROUGH THIS WHOLE FORM BEFORE YOU START ANSWERING QUESTIONS. READ EACH SECTION THEN STOP AND CARRY OUT THE INSTRUCTIONS IN TYPE LIKE THIS BEFORE YOU CARRY ON.

PILOTING OF 6-PART STORY ANALYSIS

P/
A/



Materials provided:

- Colour photocopy of client's illustrations to story
- Cassette tape of client telling story and answering questions
- Transcript of tape
- This instruction and response sheet

Instructions

Please follow these instructions given as they appear, rather than reading through this booklet completely first. Try to answer every question. Please ignore the small numbers which appear in most boxes where you will be recording your responses, they are simply for coding purposes.

Please start by filling out the following information about yourself. It will not be reported on in my writing up in any way that would identify you as an individual. Data like name and address are purely so I can keep a track of whom I have asked to do what, and not for research purposes or publication.

Name	Address	
Phone	Email	
Profession	Years in this profession	Agency (eg NHS, Social Services, voluntary, private practice)
Self-rating of experience with the 6-Part Story (please ring one number):		
1) I have had a day's training on the 6PSM but have never used it outside that context.		
2) I have tried out a 6PSM on other people, but I do not feel confident that the results I produce are valid.		
3) I have tried out a 6PSM on other people and I feel reasonably confident that I can produce some valid results via the method.		

Now spend at least three minutes looking at the illustrations drawn by the client, which has been photocopied IN COLOUR from their original drawing. **DO NOT READ THE TRANSCRIPT OR LISTEN TO THE TAPE YET.**

Now respond to the statements below. These relate just to the illustrations drawn by the client. Please note: none of the questions relate to the artistic merit or otherwise of the drawing – this is not being assessed. Tick one box only per statement, to show how far you agree or disagree with that statement.

STATEMENT	Disagree strongly	Disagree somewhat	Neutral	Agree somewhat	Agree strongly
The individual images drawn in this picture are very small.					
The drawings seem to have been done with a bold hand, making very positive, definite marks on the paper.					
Every part of this picture seems to be full of rich decoration and detail.					
The picture has obvious images of conflict, violence or death in it.					
This picture has a great many different images drawn in it.					
There is a great deal of white space left on the paper, compared with the space taken up by drawings.					
There is much use of colour made in this picture.					
This picture seems to have few recognisable elements; most of the images in it are abstract shapes or symbols.					
Just from looking at the picture, it is possible to begin to make out the story it is illustrating.					

PLEASE GO NO FURTHER UNTIL YOU HAVE LOOKED AT THE ILLUSTRATIONS AND FILLED OUT THE ABOVE BOX

Now listen to the whole of the tape recording once only. As parts of it may be unclear on first hearing, follow the transcript as you listen. (However, please do not read the transcript through before listening; use it as an aid while listening to the tape and a reminder afterwards.) Keep the picture where you can see it while you are listening to the tape.

Below is a reminder of the features of borderline, narcissistic and schizoid processes. In the space beneath each description, please note whether you

think any such features are present in the story and briefly give examples. Also indicate how strongly you think these features are present, from 0 (not present at all) to 10 (overpoweringly present.) These three ratings are independent of each other – ie they could all be 0, or all 10, or any whole number in between.

BORDERLINE	NARCISSISTIC	SCHIZOID
Themes of providing parental or idealised care, neediness or helplessness evoking rescue, being abandoned or rejected	Themes of grandiosity, perfection, admiration or specialness. Themes of humiliation, ridicule or lack of recognition.	Seeking distance from others who are threateningly close, or greater closeness to others who are disappointingly distant.
EXAMPLES:	EXAMPLES:	EXAMPLES:
STRENGTH OF FEATURES (0=absent, 10=overwhelming) <input type="checkbox"/>	STRENGTH OF FEATURES (0=absent, 10=overwhelming) <input type="checkbox"/>	STRENGTH OF FEATURES (0=absent, 10=overwhelming) <input type="checkbox"/>

PLEASE STOP AND FILL OUT THE ABOVE FORM BEFORE YOU READ FURTHER

Now re-read the transcript of the story and then respond to the statements below. Refer to the picture and replay the tape if you wish. For each statement below, indicate how far you agree with it by placing a tick in one of the five boxes available.

STATEMENT	Disagree strongly	Disagree somewhat	Neutral	Agree somewhat	Agree strongly
Magic powers, wishes or spells are important in this story.					
The story seems to be concrete, straightforward and have little hidden meaning.					
Concepts such as physical beauty or ugliness are never mentioned in this story.					
The whole atmosphere of this story is barren, bleak and lonely.					

STATEMENT	Disagree strongly	Disagree somewhat	Neutral	Agree somewhat	Agree strongly
Characters in this story seem to have belief and confidence in themselves.					
People, creatures or objects in the world are essentially helpful to the main character in completing the task.					
In this story, characters often express powerful emotions.					
The story has a clear structure of beginning, middle and end.					
The main character in this story is passive; the task is achieved by other characters intervening on their behalf.					
Solutions in this story are often about physically overcoming or destroying something.					
There are almost no relationships between active participants in this story.					
Themes of admiration or deference to a superior are very important in this story.					
It seems likely that the main character and task are good metaphors for the writer themselves and their own goals.					
There is little evidence of abstract, metaphorical thinking in this story.					
The verbal content of the story is rich, detailed and full. (Ignore the quantity and quality of drawn content.)					
At least one character in this story is sick, ill, poor or in need of rescue.					
Morbid themes of death, aggression, pain or decay predominate.					
The settings, characters and events in this story mostly have magical or fantasy elements to them.					
The elements of the story are disconnected and do not follow on from one another.					
There is an obvious bullying, dominating, violent character in this story.					
Knowledge, learning and logic seem to be important ways of coping in this story.					
Abstract concepts such as spirituality and the meaning of life are important in the story.					
This story seems likely to be a simple retelling of actual events of the teller's own life.					

STATEMENT	Disagree strongly	Disagree somewhat	Neutral	Agree somewhat	Agree strongly
The story as a whole seems to be an optimistic or positive one.					
Words describing emotional states are almost never used in this story.					
There are many examples of interactions between characters in this story.					
The storyteller seems to have been resisting the whole process of doing this story in the first place.					
Positive images of life, growth, health or production predominate.					
The main character achieves their task by their own efforts; no-one else is big or strong enough to help.					
The story as a whole seems to be pessimistic or negative.					
There is at least one character who is a powerless, overwhelmed victim in this story.					
There are many mythological creatures or themes in the story.					
Some characters in this story are tiny, empty, worthless, ridiculed or disregarded.					
There is no use of logical, problem-solving approaches by anyone in the story.					
There is a positive outcome to the story, but only through the main character's sacrifice or martyrdom.					
Themes of good and evil, right and wrong are important in this story.					
The content of the story is minimal, stark and brief.					
There is great ambivalence in this story about whether others are helpful or hurtful.					
Problems in this story tend to be about physical size and strength (or smallness/weakness.)					
There is at least one important character who is larger than life – a god, superhero, filmstar, famous historical character or similar.					
Apart from the main character, there are no other important characters in the story.					
There are several references to needles, drugs, knives or other possible tools of self-harm.					
Settings, objects and characters in the story are all recognisably from the real world.					

STATEMENT	Disagree strongly	Disagree somewhat	Neutral	Agree somewhat	Agree strongly
Themes of abandonment and being left alone by others are prominent.					
The story seems to be full of metaphor, ambiguity and potential meaning.					
In this story there is a rescuing, caring character.					
The characters in this story show no awareness of one another's needs and give one another no consideration.					
We get to know a lot about characters' emotional states in this story.					
Concepts such as physical beauty or ugliness are never mentioned in this story.					
Some characters in this story are superior, grandiose, of high rank, celebrated or admired.					

PLEASE STOP AND FILL OUT THE ABOVE FORM BEFORE YOU READ FURTHER

The next step is a formalised version of one of the methods of assessment you were taught on the 6PSM training day. Please record your thinking at each step of the way in the spaces provided. Refer to the transcript, the picture and the tape as often as you wish.

First make a list of the key actors in the story using the grid on below. This should certainly include the main character identified in the first picture; indicate which this is by underlining it. Do not forget that inanimate objects can 'act' in the story; for example a prison cell has an active role to play in confining a prisoner. Write down the list of key actors in the table below. Restrict yourself to a maximum of ten actors.

Now look again at the list of key actors and rank them in the order of importance you think they have in the story, from A (most significant) to J (least significant).

Write this ranking number down in the box to the right. The main character you have identified and underlined need not necessarily be the most significant. You may not list any actors as being of equal rank.

PLEASE STOP AND COMPLETE THE ABOVE TASK BEFORE YOU READ FURTHER

In the grid on the next page, insert the name of the most important actor (ranked A) in the row and the column marked (A). Do the same for the actors ranked B-E. Now complete the grid as you were taught on the 6PSM training day. Use these guidelines to help you.

The boxes in the grid need to be filled with active verbs. The verb needs to describe what each actor does to every other. Use the structure of a sentence to help you fill out the grid. Start with the subject of the sentence (the active role) at the top and have the object of the sentence (the passive role) at the side. Ask "What does A do to B?" Then proceed down the rows until you get down to "What does A do to E?" Then move across one column, and continue with "What does B do to A?" and so on, until you have completed the grid.

Make sure you put verbs (or verb phrases) in the grid, not nouns or adjectives. For example, if A attacks B, write "Attacks", not "Attacker" or "Attacked". As far as possible, use verbs which would be observable; for example if B fears C, what observable behaviour follows from the internal state of fear. B runs away from C? Counterattacks? Hides from? Do not use passive verbs (eg C is humiliated by B). This is better shown as an active verb in the opposite direction (B humiliates C.) Instead think of how C responds or reacts to being humiliated.

You may put more than one word in each box of the grid. You may want to use a verb phrase or write down two different but complementary verbs to retain the richness of detail in the story. Where one actor relates to another in two contradictory ways at different stages, divide the box in two so that you can separate the different actions. If this is the case where C relates to D, for example, you should also subdivide the box which shows what D does to C, because there may be a similar contradiction there.

Not every actor will have some sensible connection with every other actor. Where two actors seem to have no interaction, leave that box in the grid blank. Leave blank the shaded boxes (ie ignore what A might do to itself, B to B etc.)

		SUBJECT (Active, doing element of pair)				
		(A)	(B)	(C)	(D)	(E)
OBJECT (Passive, done-to element of pair)	(A)		90	91	92	93
	(B)	94		95	96	97
	(C)	98	99		100	101
	(D)	102	103	104		105
	(E)	106	107	108	109	

PLEASE STOP AND FILL OUT THE ABOVE FORM BEFORE YOU READ FURTHER

Now fill out the second copy of the grid below with the same five actors across top and side. Once again fill in the boxes with verbs to describe how each element relates to every other element. This time however, you have a fixed vocabulary of verbs to choose from. The 16 verbs you may use are listed below. In the definitions, S means subject (at the top of the grid) and O means object (down the side.) You may need to approximate and use a verb that is not quite the one you would have chosen. If none of the verbs seem to be appropriate, leave the box blank.

		SUBJECT (Active, doing element of pair)				
		(A)	(B)	(C)	(D)	(E)
OBJECT (Passive, done-to element of pair)	(A)		90	91	92	93
	(B)	94		95	96	97
	(C)	98	99		100	101
	(D)	102	103	104		105
	(E)	106	107	108	109	

VERB	DEFINITION
Ignores	Without giving it a second thought, S uncaringly ignores, neglects, abandons O. Without giving it a thought, S carelessly forgets O, leaves him/her out of important things.
Walls off	S walls him/herself off from O and doesn't react much. S is closed off from O and mostly stays alone in his/her own world.
Discloses to	S clearly and comfortably expresses his/her own thoughts and feelings to O. S peacefully and plainly states his/her own thoughts and feelings to O.
Affirms	S lets O speak freely, and warmly tries to understand even if they disagree. S likes O and tries to see his/her point of view, even if they disagree.

Controls	To keep things in good order, S takes charge of everything and makes O follow his/her rules. To make sure things turn out right, S tells O exactly what to do and how to do it.
Loves reactively	S relaxes, freely plays, and enjoys being with O as often as possible. S is joyful and comfortable, altogether delighted to be with O.
Attacks	Without worrying about the effect on O, S wildly, hatefully, destructively attacks him/her. Without caring what happens to O, S murderously attacks in the worst way possible.
Submits to	S thinks, does, becomes whatever O wants. S defers to O and conforms to O's wishes.
Trusts	S learns from O, relies upon O, accepts what he/she offers. S trustingly depends on O, willingly takes in what O offers.
Blames	S puts O down, blames him/her, punishes him/her. S tells O his/her ways are wrong and he/she deserves to be punished.
Protects	With much kindness, S teaches, protects and takes care of O. In a very loving way, S helps, guides, shows O how to do things.
Actively loves	S happily, gently, very lovingly approaches O and warmly invites O to be as close as he/she would like. With much love and caring, S tenderly approaches O if O seems to want it.
Emancipates	Without much worry, S leaves O free to do and be whatever O wants. Without much concern, S gives O the freedom to do things on his/her own.
Sulks towards	With much sulking and fuming, S scurries to do what O wants. S bitterly, resentfully gives in, and hurries to do what O wants.
Separates from	S knows his/her own mind and "does his/her own thing" separately from O. S has a clear sense of what he/she thinks, and chooses his/her own ways separately from O.
Recoils from	With much fear and hate, S tries to hide or get away from O. Filled with disgust and fear, S tries to disappear, break loose from O.

Now think about the story as a whole. Remind yourself of it by looking at the picture and re-reading the transcript. On the training day we discussed the possibility that the story might be seen as a set of coded messages for or requests to the therapist. If you look at this story in this way, what hidden messages might the client be sending? You may identify up to three.

Message 1: _____
 _____ 130

Message 2: _____
 _____ 131

Message 3: _____
 _____ 132

Think specifically about the task as described in the second image of the story, and the way it develops. If the main character's task could be seen as a

metaphor for the client's wish for themselves, what might they be wishing for?
(If you feel the task simply cannot be read in this way, please say so.)

Finally, about how long did it take you to complete this form (in minutes)?

If you have any feedback about this form that would help in redesigning it to make it easier to complete, please write in your comments directly at the place in question, or in the space below.

THANK YOU AGAIN!!!

NOW PLEASE RETURN THIS FORM, THE PICTURE, TRANSCRIPT AND TAPE IN THE ENVELOPE PROVIDED.

APPENDIX 2: PATIENT INFORMATION LETTER AND CONSENT FORM

11 November 2002

My Ref: S074

Dear Sir/Madam,

RESEARCH INTO THE SIX-PART STORY: CLIENT INFORMATION SHEET AND CONSENT FORM

I am writing to ask if you would be willing to help me in some research I am undertaking at the moment. Please take the time to read this information carefully. Do discuss it with friends, relatives, your key worker and anyone else you would like to advise you.

WHAT'S THE POINT OF THE RESEARCH?

I am studying an assessment called the 6-Part Story Method (6PSM). This involves a client and their worker creating a short, fictional story together, which they then discuss. We believe that this can be a very good way of finding out new information and improving the client-worker relationship.

WHAT AM I ASKING YOU TO DO?

I would like to meet you and spend about an hour creating a 6-Part Story together. This will be tape recorded and typed up later. During that session I will also give you a questionnaire to fill out, which will help me in the next step. You and I will then meet for a second time and I will conduct an interview with you that will focus on your present difficulties, personality style and so on. At the end of the interview I will also ask you to fill out two more short questionnaires. One month after doing the first 6-Part Story, you and I will meet for a third time and create a second story together, but I will not need to do the interview and questionnaires again.

WHY HAVE YOU BEEN CHOSEN?

You have been randomly chosen as a representative of the clients who use community mental health services and because your key worker thinks this process might be helpful in their work with you.

DO YOU HAVE TO TAKE PART?

Certainly not! Your participation is entirely voluntary. If you decide not to take part then the existing care you receive from your key worker will not change in any way as a result. You can agree now and change your mind later. If you do this I will remove your story from the research.

WHAT WILL HAPPEN TO THE MATERIAL PRODUCED?

I will discuss with your key worker the 6-Part Story and the results of the questionnaires you filled out. I will give him or her a written summary of these which I am very happy

for you to see, along with a transcript of your story sessions with me. These will be kept in your confidential casenotes. I will take copies of your questionnaire answers, your 6-Part Story, the tape recording and the transcript of the tape. Your name and any other identifying information will be removed from these before they leave NHS premises. As part of my research I will play your tape recording to a small number of other clinicians who are helping me in the research. They will not know you or be able to identify you by anything they see or hear.

WILL THERE BE ANY BENEFITS TO AGREEING?

I hope that the 6-Part Story will be useful to you and your key worker. It may help you to open up and talk to your worker in a way you have not previously been able to do. They may find themselves able to understand you better than they could before. Your key worker will also get the results of the questionnaires and interview that I do, which may help them in understanding you better and providing you with a care plan that suits you better.

WHAT WILL HAPPEN TO THE RESULTS OF THE RESEARCH STUDY?

I will write the results up in a thesis which I will go into the Hull University library. I will also write shorter articles for publication in scientific journals. If you would like I will send you a copy of any article I write.

WHAT ABOUT CONFIDENTIALITY?

All information about you collected during this research will be kept strictly confidential. Any information about you that leaves the NHS setting will have your name, address and any other identifying features removed. I will let your own GP and your psychiatrist (if you have one) know that you have agreed to let me use your story and other information. Neither you nor any individual will be identified in the thesis or any article.

WHO IS ORGANISING, FUNDING AND REVIEWING THE RESEARCH?

The research is being funded by the NHS Regional Executive for the Northern and Yorkshire region. It is being supervised by the Clinical Psychology Department at Hull University. The research has been approved by the Hull and East Riding Local Research Ethics Committee.

WHO DO I CONTACT FOR FURTHER INFORMATION?

In the first instance please contact your key worker if you have any further questions, either now or in the future.

Yours sincerely,

KIM DENT-BROWN

RESEARCH FELLOW IN DRAMATHERAPY

Ref No: _____

CLIENT CONSENT FORM
RESEARCH IN THE SIX-PART STORY METHOD

Researcher: Kim Dent-Brown
Specialist Therapies Service
Miranda House
Gladstone Street
HULL HU3 2RT
Phone (01482) 617503

- 1) I confirm that I have read and understood the information in the information sheet 11 November 2002 for the above study and have had the opportunity to ask questions.
- 2) I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my care or legal rights being affected.
- 3) I understand that sections of my medical notes may be looked at by the University and local Research Ethics authorities who regulate the research where it is relevant. I give permission for these individuals to have access to my records.
- 4) I give permission for the researcher to contact my GP (and Psychiatrist if I have one) and let them know that I am taking part in this research.
- 5) I agree that:
 - The researcher may use my story, drawings, questionnaire and interview responses and tape recordings in research.
 - The researcher may keep anonymised copies of the above material during the period of the research.
 - The researcher may allow other professionals to have access to these materials to gather further data during his research.

Name of Client	Date	Signature

Name of Researcher	Date	Signature

1 copy to client, 1 copy to researcher, 1 copy to client's medical notes.

APPENDIX 3: PROTOCOL FOR ELICITING 6-PART STORY MATERIAL

RESEARCH INTO THE SIX-PART STORY METHOD

SIX-PART STORY SCRIPT FOR CLINICIANS

Throughout this document, text in italic like this is instructions to you. Text in type like this is the words you will actually say to the client. As far as possible, ask all the questions and give all the instructions that are presented here, but do not ask any further questions or give further information until after the tape is switched off.

Questions or instructions that are indented like this are only to be asked in certain circumstances. It will be clear from the context whether you should ask them or not.

SOME PROBLEM-SOLVING STRATEGIES:

*If the client asks if they can write words as well as draw pictures, say: **Try not to use words, but find an image, picture or symbol that will remind you of the thing you're trying to illustrate. But don't worry if they insist on writing something down.***

If a client gets stuck or needs encouragement, use any of the following types of prompt:

- **It doesn't matter how odd or silly it seems, just say/draw the first thing that comes into your head.**
- **If you don't know what's happening, just make it up. You're in charge of the story and can't get it wrong.**
- **Can you tell me a bit more about.....**
- **If you don't know the answer, make it up on the spot!**
- **There's no wrong way of doing this; whatever you do will be right.**
- **Is there anything else important about....**
- **If we were actually there, what would we be aware of?**
- **If this was being shown as a film, what else would we be seeing on screen?**

If the teller shifts during the telling from the third person (he, she, it, they) to the first person (I, we), encourage them to return to the third person form. If they persist, leave them to it.

If a client asks whether they are doing it right, tell them they are. If they ask for clarification of an instruction, repeat the instruction or paraphrase it without adding to it.

Please give clients every encouragement to complete the process, but if they make it clear they want to stop, respect this. Similarly, if your professional judgement is that you should stop the process or depart from this script in the client's best interests, then you must follow your judgement.

The script starts below. Good luck!

PART ONE – DRAWING THE STORY

This activity is a way of both of us finding out more about the way you see yourself, others around you and the world. It's a bit different from the straightforward "question and answer" approach you might be used to. I'm going to help you to create and then tell me a story, which we will then discuss together. Don't worry if you feel you're no good at creating things; I'll take you through the process step by step. The first step is very simple; I have a piece of paper and some pens here – please divide the paper into six spaces for you to work in. You are going to make a story in six parts, and need six spaces on the paper, one for each part.

[Give client blank white A4 paper on clipboard and small felt-tip pens (not flipchart size) in red, orange, yellow, green, blue, purple, brown and black.]

There are just a couple of rules to follow which seem to make this task more successful. One is that you are going to make up a new story; not just retell me "Cinderella" or "My first day at school". I'd like you to make up a brand new, completely fictional story that's never been told before. The second rule is that this seems to work best if you set it as far away from your own real life as possible. For example, if you were a librarian it be a bad idea to set this story in a library, or have someone like yourself as the main character. Your story can be in any other style you like; a fairytale, a myth, a historical drama, a soap opera, science fiction, a kids' cartoon..... If you do set it in the real world, use settings and characters that you don't know yourself, but which you have to imagine. You're going to draw each part of the story first, following my instructions. When it's all drawn you're going to tell it to me and afterwards I'll ask you some questions about it. You don't need to write anything down – you will be doing it all with pictures.

Now in your first space I want you to draw a picture or sketch of the main character of your story. This need not be human – it could be an animal or a fantasy character or anything that you could imagine might be able to talk and think and feel. Don't forget, make this main character someone or something as unlike you as possible.

Pause for them to start drawing. If they have trouble getting started, you can say: Now what often happens is that you have lots of things in mind and say to yourself "No I can't put that, it's stupid." But a good idea is always to just to stick with whatever first comes to mind – it doesn't matter whether it makes sense to you or not.....

If they still can't seem to get going, you could say: If you're still stuck, you might find it easier to start with a setting and sketch that in first. For example, does this story have a setting indoors or outdoors? In some fantasy world or this one? If this one, is it historical, present day or future? Once you have a setting, you can think what kind of character might appear here and sketch them in.

Once you have your main character, draw in the setting behind them, still in the first space, so we can see where they appear, what kind of place we find them.

Pause at this point to let the client draw their first picture. When they seem to have finished, ask them if it is all right to go on to the next part. Allow similar time after each part that follows.

In the second space I'd like you to show us what the main character's task is. At the beginning of this story they are faced with some task, some thing they have to do by the end of the story. It could be a journey to be made, an enemy to be defeated, a lost object to be found, an obstacle to be overcome; anything at all.

If they haven't started drawing yet, you can say: If you have trouble thinking of something, use the information in the first picture to help you. What kind of task would this kind of character, in this kind of setting, be faced with? It doesn't matter if the task seems very big, or very small, or whether it makes sense or seems silly. Just try and find a way of representing the task in a drawing in the second space.

Pause to allow second drawing.

The third space is where it starts to become a story. If this task was simple, and the character just did it, the story would soon be over. What makes this a story is that there are some things which make it harder for the main character. Now these might fall into roughly three types. There might be inabilities of the main character – for example if your story was about a fish who had to go up a mountain, then the fact the fish doesn't have legs would be a problem. Or there might be obstacles or problems in the world; a fence that has to be crossed or a locked door. Or there might be other people or creatures who actively try to stop your main character and make things more difficult for them. So in this third space there might be two or three different images of things which cause a problem to your main character in achieving their task.

Pause to allow third drawing.

The fourth space is rather like the opposite of the third, because there are some things which will help your main character. These might be strengths or abilities of your main character – perhaps my fish from the example earlier could be a flying fish. Or there might be things you character finds in the world – tools or other objects which might be useful. Or there might be other people or creatures who actively help or advise your main character. Once again, you might have two or three different images in this space.

Pause to allow fourth drawing.

The fifth space is for the main action of the story. This is where all the four parts you have drawn so far come together, and we see what happens. It is where we see whether – and how – your main character succeeds (or maybe fails) in their task. Or maybe by this point in the story the task has changed; that's not a problem. There may be a lengthy piece of main action, with a whole sequence of things happening. You may need to just pick the crucial moment, the turning point, the decisive instant that determines success or failure and illustrate that.

Pause to allow fifth drawing.

The sixth space is for what happens after the main action. It's like the last scene in a film or the last chapter of a book – the main action is over and we see what follows from it. How is the world different now that the main character has succeeded or failed? Maybe it isn't different at all? In a traditional fairytale the ending is often 'And they all lived happily ever after'. Is this what happens in your story or does something else happen. It doesn't matter if the ending seems untidy or unfinished; this may be as much the start of a new story as the end of this one.

PART TWO – TELLING THE STORY

From this point onwards the tape should be switched on.

Now I'm going to ask you to tell me the whole story through from the beginning, using the pictures to help you through. I won't interrupt you or ask you any questions, unless you get stuck and can't remember what each part was supposed to be. I'd like you to try and add as much detail as you can in the telling, so don't just rattle through it quickly. For example, rather than saying 'Once upon a time there was a fish who had to climb a mountain' invent as much as you can about each bit of the story. Give me more detail about every character and place in the story. Once you've gone through the story I'll go back and ask you some questions about each of the pictures. So tell me the story through now. A good way to start is by saying 'Once upon a time there was a.....'.

NB: If your teller forgets which picture was what, the six parts in order are:

(1) Main character and setting, (2) Task, (3) Things that hinder, (4) Things that help, (5) Main action, (6) What happens afterwards.

PART THREE – ASKING QUESTIONS

When asking the following questions, just ask the main question first. If you get at least a couple of sentences with some new detail, that's enough. Go on to the next main question. If the teller is struggling, ask the inset supplementary questions to help them.]

Now that you've told me the story once through, can I ask you some more questions in detail about each of the pictures one by one?

What about the first picture. You said your main character was a [repeat

details] Can you tell me a bit more about them?

Only if answer is very short:

Can you tell me a bit more; if the main character was someone you knew well and you were giving me a description. How else would you describe [main character]? Give me some other descriptions of their personality, of what they're like.

If detail still lacking or client doesn't know where to start:

How is [main character] feeling at the start of the story? What mood is [he, she, it] in at the beginning?

Staying in the first picture, tell me about the place a bit more. Tell me a bit more about where we find the main character; what kind of place is it?

If prompting needed: **Just say a bit more about the location; what it feels like, what kind of atmosphere it has, what we'd see or hear if we were there right now.**

The second picture was of the task; can you tell me a bit more about the task? What it involves, why it's important?

Only if answer is very short: **Who gave [main character] their task? Was it someone else, or did they set it themselves or what?**

If detail still lacking or client doesn't know where to start: **How does [main character] feel about having to do [task]? What's their reaction, what do they think about having this job to do?**

What would happen if [character] simply refused to do the task and said "No, I'm not doing that". Or if they tried and failed, what would happen then?

So in the third picture we have the things that make it harder for the main character. Just tell me a bit more about the things that make it difficult.

Only if answer is very short: **How does each of these things make it difficult for [main character] to achieve their task? How do they get in the way exactly?**

[If detail still lacking or client doesn't know where to start:] **If the [hindering factors] had a voice and could talk to [main character], what would they be saying? And what would [main character] say back to them?**

How does [main character] think and feel about the things that make it difficult? What thoughts or emotions do they have when things get in the way?

So in the fourth picture, remind me of the things that help [main character]. Who or what are they and how do they help exactly?

Only if answer is very short: **If the [helpful factors] had a voice and**

could talk to [main character], what would they be saying? And what would [main character] say back to them?

How does [main character] respond to being helped? How does [he, she, it] react towards the things that help?

Tell me again what's happening in the fifth picture, the main action. Can you give me some more detail about what's going on here?

[Only if answer is very short:] **What's [main character] thinking and feeling as all this goes on? What state are they in as the main action unfolds?**

So this fifth picture shows how the main action actually went; how else might it have gone? If there was more or less help or hindrance, might things have turned out differently?

So, remind me what's going on in the final picture. Can you explain it a bit more and give me some more detail?

[Only if answer is very short:] **So what's [main character] thinking and feeling now we've got to the end of the story? What state are they in, what mood?**

Looking right back to the beginning, and comparing it with where we are now at the end of the story, what do you notice? What's different and what's the same?

I've asked lots of questions about the story – is there anything I haven't asked about, or that you haven't had chance to talk about? Is there anything that didn't seem important at first but now seems more so?

If this story as a whole had a title, what title would you give it?

Repeat title for confirmation.

And what if this story was like one of those parables or fables, that as well as being a story has some kind of teaching in it. What lesson, moral or advice does the story have for you?

In the FIRST session you have with the client, please ask the following two questions to finish:

Now we're nearly at the end of the questions, I just wondered how you have found the whole process. What's it been like doing this storymaking exercise, what are your first reactions to it?

Now that you've done the story, do you think it tells us anything about you? Are bits of you in there somewhere?

Please finish the first session by giving the client the SCID-II questionnaire and telling them that I will be arranging to meet them soon. They can give the

questionnaire back to me then. You should also remind them that you will be doing a second 6-part story in a month's time.

In the SECOND session you have with the client, please ask these two questions to finish:

What has it been like doing this again a second time? Are there any things that seem very similar or very different to last time?

Do you think this exercise has been useful to you and I for our work, or a bit of a waste of time? And what is it that was good or bad about it?

*If the client's reaction is mostly positive, please ask: **So if you think it works, HOW does it work? What do you think is happening when you tell the story and we discuss it?***

*If the client's reaction is neutral or negative, please ask: **Is there anything about it that was particularly unhelpful, or that you think should be done differently?***

Please finish the second session by reminding the client that I will be sending the tape, transcript and pictures back to you once I have copied them. I will also be sending a short report for you and the client, based on the stories, my interview and the questionnaires.

APPENDIX 4: INSTRUCTIONS TO FINAL RATERS AND FINAL RATING FORM

To: Raters in the
6-Part Story Method
(6PSM) research project

21 January 2003

Dear Colleague,

PLEASE READ THIS LETTER BEFORE YOU OPEN ANY OF THE ENVELOPES

Many thanks again for helping me in this research. The eight envelopes enclosed each contain:

- A colour photocopy of the storyteller's drawing
- A transcript of their storytelling session
- A rating form for you to fill in

Please work through the envelopes in numerical order. Please do not open all the envelopes at once; open the first one and complete the rating form before moving on to number two.

The rating form gives you all the instructions you need; in outline, the first section (3 questions) of the form is about the picture, and the second and third sections are about the story as a whole. The fourth and last section asks you to make some guesses about the person who wrote the story.

Please try to respond to every question. If you have difficulty understanding or responding to the questions please contact me by phone or email and I will try and help you out. The first story you go through may take up to an hour, but the others should go more quickly once you are used to the format, probably no more than 20-30 minutes per story and maybe less than that.

Please do not make copies of the stories or pictures, as the participants have not given permission for that. Once you have finished all eight stories, please return them to me in the envelope enclosed. If you do not manage to finish, please return all the stories to me anyway.

Good luck and thanks again!

Yours sincerely,

KIM DENT-BROWN
RESEARCH FELLOW

RATER NO:	
TAPE NO:	

Disagree strongly	Disagree somewhat	Neutral	Agree somewhat	Agree strongly
----------------------	----------------------	---------	-------------------	-------------------

SECTION 1: FIRST LOOK AT THE PICTURE AND RESPOND TO THESE THREE STATEMENTS BEFORE READING THE STORY

1	Every part of this picture seems to be full of rich decoration and detail.						
2	The picture has obvious images of conflict, violence or death in it.						
3	There is much use of colour made in this picture.						

SECTION 2: NOW READ THE STORY AND GO ON THE RESPOND TO THE REST OF THE STATEMENTS

CHARACTERS IN THE STORY

4	In this story there is a rescuing, caring character.						
5	The main character is specifically described as average, ordinary, normal.						
6	The main character has likeable, admirable qualities						
7	Some characters in this story are superior, grandiose, of high rank, celebrated or admired.						
8	There is at least one important character who is larger than life – a god, superhero, filmstar, famous historical character or similar.						
9	There is at least one character who is a powerless, overwhelmed victim in this story.						
10	There is an obvious bullying, dominating, violent character in this story.						
11	At least one character in this story is sick, ill, poor or in need of rescue.						
12	Characters in this story seem to have belief and confidence in themselves.						

THE MAIN CHARACTER'S TASK

13	The main task is of self-development, satisfying curiosity, enlightenment, achieving potential, or a spiritual journey.						
14	The main task is about a journey or escape from a hostile or unpleasant situation.						
15	Failure to complete the task would mean starvation or death for the main character.						
16	The main task is to capture, kill or eat some other creature or being.						

DEALING WITH PROBLEMS

17	Violent, physical means are used to overcome obstacles.						
18	Opposition makes the main character angry, frustrated...						
19	Problems in this story tend to be about physical size and strength (or smallness/weakness.)						
20	Obstacles are overcome with faith, persuasion, skill or problem-solving.						
21	The main character heroically overcomes stronger opponents.						

		Disagree strongly	Disagree somewhat	Neutral	Agree somewhat	Agree strongly
RELATIONSHIPS IN THE STORY						
22	The characters in this story show no awareness of one another's needs and give one another no consideration.					
23	Others are mostly seen as helpful, positive, friendly.					
24	Others are mainly a threat towards the main character in the story.					
25	There are many examples of interactions between characters in this story.					
26	Themes of abandonment and being left alone by others are prominent.					
27	Other characters essentially complete the task for the main character.					
28	Others are helpful because their presence means the main character is not alone.					
THE FINAL OUTCOME						
29	The outcome is a 'win-win' situation for main character and most others.					
30	The outcome is positive for the main character.					
31	The outcome is positive for the main character, but only at the expense of most others.					
32	The outcome is negative for the main character.					
MORAL OR ADVICE OF THE STORY						
33	The story teaches that confronting fears and problems is necessary to happiness.					
34	The moral is that we all have untapped potential in us.					
35	The moral of the story is about persistence, brains or bravery achieving success.					
36	The story teaches that we always need other people.					
GENERAL IMPRESSIONS						
37	The story as a whole seems to be pessimistic or negative.					
38	Positive images of life, growth, health or production predominate.					
39	Magic powers, wishes or spells are important in this story.					
40	There are many mythological creatures or themes in the story.					
41	The whole atmosphere of this story is barren, bleak and lonely.					
42	Morbid themes of death, aggression, pain or decay predominate.					
43	The settings, characters and events in this story mostly have magical or fantasy elements to them.					
44	Themes of good and evil, right and wrong are important in this story.					
45	The content of the story is minimal, stark and brief.					
46	The story as a whole seems to be an optimistic or positive one.					
47	Things that harm others are important to the story – guns, swords, teeth, claws.					

SECTION THREE. Think about the story as a whole. Look at the list of 8 wishes below, and put a 1 next to the one you think best represents the main character's wish in the story. Put a 2 next to another wish, if it also describes the main character's wishes well. Do not check more than two boxes.

WISH OF MAIN CHARACTER

1	The main character wants to assert themselves and be independent, have self-control, be their own person.	
2	The main character wants to oppose, hurt or control others.	
3	The main character wants other people to be responsible for them; to control them, help them or hurt them.	
4	The main character wants to be distant and avoid conflicts with others, to not be hurt by others.	
5	The main character wants to be close and accepting, to respect others, to be open to others and have others open up to them.	
6	The main character wants to be loved and understood, liked, accepted and respected by others.	
7	The main character wants to feel good and comfortable, to have stability, to feel happy and good about themselves.	
8	The main character wants to achieve things and help others; to better themselves, to be good, to be useful.	

Now think about how other elements (people, animals, things...) in the story respond to the main character and their wish. Look at the list of 8 responses below, and put a 1 next to the one you think best represents the response of other elements in the story. Put a 2 next to another response, if it also describes well how other elements respond to the main character. Do not check more than two boxes.

RESPONSE OF OTHERS TO MAIN CHARACTER'S WISH

9	Others are strong, independent and happy.	
10	Others are strict and controlling.	
11	Others get upset. They are hurt, dependent, anxious, angry or out of control.	
12	Others are bad and cannot be trusted.	
13	Others reject and oppose the main character; they do not trust or understand the main character, they are disrespectful, unhelpful or hurtful.	
14	Others are helpful and co-operative towards the main character.	
15	Others like the main character, love them, give them independence.	
16	Others are understanding, open and accepting towards the main character.	

Now think about how the main character responds to other elements in the story. Look at the list of 8 responses below, and put a 1 next to the one you think best represents the response of the character to other elements in the story. Put a 2 next to another response, if it also describes well how the main character responds to other elements. Do not check more than two boxes.

RESPONSE OF MAIN CHARACTER TO OTHERS

17	The main character is open, understanding and helpful.	
18	The main character does not understand or like others, is not open and receptive towards others.	
19	The main character feels respected and accepted, feels comfortable, happy and loved, feels as if they are like others.	
20	The main character opposes other people and hurts them.	
21	The main character feels self-confident and independent, self-controlled and in control.	
22	The main character feels helpless, out of control, uncertain and dependent.	
23	The main character feels disappointed and depressed, angry, unloved and jealous.	
24	The main character feels anxious, ashamed and guilty.	

SECTION FOUR - TWO FINAL MISCELLANEOUS QUESTIONS

What is your best guess about the mental health of the story's author? Write down a number between 1-10, where: 1=poorest imaginable mental health and 10= best imaginable mental health.		Make a guess at the gender of the person who wrote this story. Are they male or female?	
---	--	---	--

APPENDIX 5: RESULTS OF MINOR ANALYSES (SEE SECTION 10.8)

A number of characteristics of stories and tellers were compared to one another, using a chi-square test in the case of categorical data, a Spearman's correlation in the case of scale data and a t-test in the case of mixed (scale vs categorical) data. The characteristics tested were:

- Gender of teller (categorical)
- Gender of main character (categorical)
- Type of main character (categorical)
- Inferred mental health (scale)
- LIWC-B score (scale)
- Pessimism failure score (scale)
- Education of teller (categorical)

Those results that have not been reported are briefly summarised below. Significant associations included:

- Stories with humanoid main characters had a significantly lower LIWC-Borderline score than stories with other types of character ($t = 3.27$, $df = 55.9$, $p < .01$, equal variances not assumed).
- Authors who continued in school after 16 were more likely to choose a main character of the opposite gender than those who had left school earlier (chi-squared $\chi^2 = 5.11$, $df = 1$, $p < .05$). This became even more significant when clinicians (who could all be safely assumed to have continued education beyond 16) were added in to the analysis ($\chi^2 = 11.72$, $df = 1$, $p < .01$).

While there was no significant association in the cases of:

- Gender of teller vs Type of main character: No significant association ($\chi^2 = 1.65$, $df = 4$, $p > .05$)
- Gender of teller vs Inferred mental health: No significant difference ($t = -1.62$, $df = 63$, $p > .05$).
- Gender of teller vs LIWC-B score: No significant difference ($t = 0.04$, $df = 63$, $p > .05$).
- Gender of teller vs Negativity score: No significant difference ($t = 0.53$, $df = 62$, $p > .05$).
- Gender of main character vs Type of main character: No significant association ($\chi^2 = 9.34$, $df = 8$, $p > .05$).
- Gender of main character vs Inferred mental health: No significant difference ($t = 0.57$, $df = 61$, $p > .05$).
- Gender of main character vs LIWC-B score: No significant difference ($t = -0.05$, $df = 61$, $p > .05$).
- Gender of main character vs Negativity score: No significant difference ($t = -.09$, $df = 60$, $p > .05$).
- Type of main character vs Inferred mental health: No significant difference ($t = -.77$, $df = 63$, $p > .05$).
- Type of main character vs Negativity score: No significant difference ($t = 0.85$, $df = 62$, $p > .05$).

- Apart from the association with choice of main character gender mentioned above, there was no significant association between length of education and any other of these variables.

18.0 BIBLIOGRAPHY

- Abma, T. A. (1998). Storytelling as inquiry in a mental hospital. *Qual Health Res*, 8(6), 821-838.
- Ackerman, S. J., Clemence, A. J., Weatherill, R., & Hilsenroth, M. J. (1999). Use of the TAT in the assessment of DSM-IV cluster B personality disorders. *Journal of Personality Assessment*, 73(3), 422-448.
- American Psychiatric Association. (1980). *The diagnostic and statistical manual of psychiatric disorders - III*. Washington, DC: Author.
- American Psychiatric Association. (1994). *The diagnostic and statistical manual of psychiatric disorders - IV*. Washington, DC: Author.
- Andersen-Warren, M., & Grainger, R. (2000). *Practical approaches to dramatherapy: The mask of Perseus*. London: Jessica Kingsley.
- Aristotle. (350BC). *Poetics - Part VI*. Trans S.H. Butcher. Retrieved 12 January, 2004, from the World Wide Web:
- Arntz, A. (1999). Do personality disorders exist? On the validity of the concept and its cognitive-behavioural formulation and treatment. *Behaviour Research and Therapy*, 37, S97-S134.
- Arthern, J., & Madill, A. (1999). How do transitional objects work?: The therapist's view. *British Journal of Medical Psychology*, 72(1), 1-21.
- Arthern, J., & Madill, A. (2002). How do transitional objects work?: The client's view. *Psychotherapy Research*, 12(3), 369-388.
- Barham, M. (2003). Practitioner-based research: Paradigm or paradox? *Dramatherapy*, 25(2), 4-7.
- Barkham, M., Hardy, G. E., & Startup, M. (1996). The IIP-32: A short version of the Inventory of Interpersonal Problems. *British Journal of Clinical Psychology*, 35, 21-35.
- Bengston, D., & Xu, Z. (1995). *Changing national forest values: A content analysis (Research paper NC, 323)*. United States Department of Agriculture, Forest Service, North Central Forest Experiment Station, St. Paul, MN. Retrieved 27 August, 2002, from the World Wide Web: <http://www.ncrs.fs.fed.us/searchtools/viewpub.asp?key=806>
- Benjamin, L. S. (1996). *Interpersonal diagnosis and treatment of personality disorders* (2nd ed.). New York: The Guilford Press.
- Bland, J. M., & Altman, D. G. (1995). Statistics notes: Multiple significance tests: the Bonferroni method. *BMJ*, 310(6973), 170-.
- Burgmer, M., Jessen, F., & Freyberger, H. (2000). Polythetic diagnostic approach to the borderline personality disorder - The valency of the

- single criterion in the concept of professional therapists.
Psychopathology, 33(3), 119-124.
- Castillo, H. (2001). *Grounds for optimism with Personality Disorders*. Paper presented at the Third Annual conference of the James Nayler Foundation, London, available on the www at [http://www.jnf.org.uk/main/4confs/Earlier_conferences/2001.htm# Heather Castillo](http://www.jnf.org.uk/main/4confs/Earlier_conferences/2001.htm#_Heather_Castillo).
- CORE System Group. (1999). *CORE system user manual*. Leeds: CSG.
- Costantino, G., & Malgady, R. G. (1999). The Tell-Me-A-Story Test: A multicultural offspring of the Thematic Apperception Test. In L. Gieser & M. I. Stein (Eds.), *Evocative images: the Thematic Apperception Test and the art of projection* (pp. 191-206). Washington, DC: American Psychological Association.
- Coulacoglou, C. (1995). Personality assessment using the Fairy Tale Test (FTT), the H-T-P and the TAT: A case study. *British Journal of Projective Psychology*, 40(1), 1-12.
- Coulacoglou, C., & Kine, P. (1995). The Fairy Tale Test: A novel approach in projective assessment. *British Journal of Projective Psychology*, 40(2), 10.
- Creswell, J. W. (1994). *Research design: Qualitative & quantitative approaches*. Thousand Oaks, California: Sage.
- Crimmens, P. (1998). *Storymaking and creative groupwork with older people*. Bristol, PA, US: Jessica Kingsley.
- Dent-Brown, K. (1999a). The six-part story method (6PSM) as an aid in the assessment of personality disorder. *Dramatherapy*, 21(2), 10-14.
- Dent-Brown, K. (1999b). Working with personality disorder in community mental health. *Dialogue*, 2, 5.
- Dent-Brown, K. (2001a). The chance of a lifetime: A dramatherapist visits the Middle East. *Dramatherapy*, 23(2), 26-29.
- Dent-Brown, K. (2001b). Story as therapeutic tool: The Six-Part Story Method. *Context*, 55, 22-23.
- Dey, I. (1993). *Qualitative data analysis: A user-friendly guide for social scientists*. London: Routledge.
- Dickson, D. H., & Kelly, I. W. (1985). The Barnum effect in personality assessment - a review of the literature. *Psychological Reports*, 57(2), 367-382.
- Driessen, M., Beblo, T., Reddemann, L., Rau, H., Lange, W., Silva, A., Berea, R. C., Wulff, H., & Ratzka, S. (2002). Ist die Borderline-Persönlichkeitsstörung eine komplexe posttraumatische Störung? Zum

- Stand der Forschung [Is Borderline Personality Disorder a complex post-traumatic disorder? - The state of research]. *Nervenarzt*, 73(9), 820-829.
- Dunn, M., & Parry, G. (1997). A formulated care plan approach to caring for people with borderline personality disorder in a community mental health service setting. *Clinical Psychology Forum*, 104(June), 19-22.
- Dwivedi, K. N. (Ed.). (1997). *The therapeutic use of story*. London: Routledge.
- Edwards, J. (1996). Examining the clinical utility of the Moreno Social Atom Projective Test. *Journal of Group Psychotherapy, Psychodrama and Sociometry*, 49(2), 51-75.
- Exner, J. E., Jr. (1995). *The Rorschach: A comprehensive system* (3rd ed. Vol. 1: Basic Foundations). New York: John Wiley and Sons.
- Feyerabend, P. (1975). *Against method*. London: Verso.
- Feyerabend, P. (1978). *Science in a free society*. London: Verso.
- First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B. W., & Benjamin, L. S. (1997). *User's guide for the Structured Clinical Interview for DSM-IV axis I personality disorders*. Washington, DC: American Psychiatric Press.
- Flanagan, R. (2000). Objective and projective personality assessment: The TEMAS and the Behavior Assessment System for Children, self-report of personality. *Psychological Reports*, 84, 865-867.
- Flanagan, R., & di Giuseppe, R. (1999). Critical review of the TEMAS: A step within the development of thematic apperception instruments. *Psychology in the Schools*, 36(1), 21-30.
- Forrester, A. M. (2000). Role-playing and dramatic improvisation as an assessment tool. *The Arts in Psychotherapy*, 27(4), 235-243.
- Frank, L. K. (1939). Projective methods for the study of personality. *The Journal of Psychology*, 8, 389-413.
- Freud, S. (1911). On the mechanism of paranoia. In S. Freud (Ed.), *General psychological theory: papers on metapsychology*. New York: Collier.
- Gersie, A. (1991). *Storymaking in bereavement*. London: Jessica Kingsley.
- Gersie, A. (1992). *Earthtales: storytelling in times of change*. London: Green Print.
- Gersie, A. (1997). *Reflections on therapeutic storymaking: The use of stories in groups*. London: Jessica Kingsley.
- Gersie, A. (2002). Some thoughts about the development of Therapeutic Storymaking. *The Prompt (The magazine of the British Association of Dramatherapists)*, Winter 2002/03, 6-7.

- Gersie, A. (2003a). Further thoughts about question-based storymaking. *The Prompt (The magazine of the British Association of Dramatherapists)*, Summer 2003, 6-8.
- Gersie, A. (2003b). Some thoughts about the question-based storymaking technique. *The Prompt (The magazine of the British Association of Dramatherapists)*, Spring 2003, 6-7.
- Gersie, A., & King, N. (1990). *Storymaking in education and therapy*. London: Jessica Kingsley.
- Glaser, B., & Strauss, A. (1967). *Discovery of Grounded Theory*. Chicago: Aldine.
- Gottschalk, L. A., Stein, M. K., & Shapiro, D. H. (1997). The application of computerized content analysis of speech to the diagnostic process in a psychiatric outpatient clinic. *Journal of Clinical Psychology*, 53(5), 427-441.
- Grainger, R. (1999). *Researching the arts therapies: A dramatherapist's perspective*. London: Jessica Kingsley.
- Grainger, R. (2002). Thoughts on the future of dramatherapy research. *Dramatherapy*, 23(3), 9-11.
- Greimas, A. J. (1966). *Sémantique structurale : Recherche de méthode*. Paris: Larousse.
- Hanson, W. E., Claiborn, C. D., & Kerr, B. (1997). Differential effects of two test-interpretation styles in counseling: A field study. *Journal of Counseling Psychology*, 44(4), 400-405.
- Holmes, D. S. (1974). The conscious control of thematic projection. *Journal of Consulting and Clinical Psychology*, 42, 323-329.
- Holmstrom, R. W., Karp, S. A., & Silber, D. E. (1994). Prediction of depression with the Apperceptive Personality Test. *J Clin Psychol*, 50(2), 234-237.
- Holt, R. R. (1992). Formal aspects of the TAT: A neglected resource. In E. I. Megargee & C. D. Spielberger (Eds.), *Personality assessment in America: A retrospective on the fiftieth anniversary of the Society for Personality Assessment*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureno, G., & Villasenor, V. S. (1988). Inventory of Interpersonal Problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology*, 56, 885-892.
- Huprich, S. K. (2001). Object loss and object relations in depressive personality analogues. *Bull Menninger Clin*, 65(4), 549-559.

- Jackson, H. J., Gazis, J., Rudd, R. P., & Edwards, J. (1991). Concordance between two personality disorder instruments with psychiatric inpatients. *Compr Psychiatry*, 32(3), 252-260.
- Jacobsberg, L., Perry, S., & Frances, A. (1995). Diagnostic agreement between the SCID-II screening questionnaire and the Personality Disorder Examination. *J Pers Assess*, 65(3), 428-433.
- Jensen, A. R. (1965). A review of the Rorschach. In O. K. Buros (Ed.), *Sixth mental measurements handbook*. Highland Park, New Hampshire: Gryphon.
- Johnson, D. R., & Quinlan, D. M. (1993). Can the mental representations of paranoid schizophrenics be differentiated from those of normals? *Journal of Personality Assessment*, 60(3), 588-601.
- Jones, P. (1993). Dramatherapy: Five core processes. *Dramatherapy*, 14(1), 8-15.
- Keown, P., Holloway, F., & Kuipers, E. (2002). The prevalence of personality disorders, psychotic disorders and affective disorders amongst the patients seen by a community mental health team in London. *Social Psychiatry & Psychiatric Epidemiology*, 37(5), 225-229.
- Kernberg, O. (1967). Borderline personality organisation. *Journal of the American Psychoanalytic Association*, 15, 641-685.
- Kernberg, O. (1984). *Severe personality disorders: Psychotherapeutic strategies*. New Haven/London: Yale University Press.
- Kleindienst, M., & Frude, N. (1999). *A review of research relevant to outcome measurement in the arts therapies*. Cardiff: University of Wales College, Department of Psychology.
- Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Beverly Hills: Sage.
- Lahad, M. (1992). Storymaking in assessment method for coping with stress. In S. Jennings (Ed.), *Dramatherapy theory and practice II* (pp. 150-163). London: Routledge.
- Lahad, M., & Ayalon, O. (1993). BASIC Ph - The story of coping resources, *Community stress prevention* (Vol. II). Kiryat Shmona, Israel: Community Stress Prevention Centre.
- Leach, C., Lucock, M., Barkham, M., Noble, R., & Iveson, S. (in preparation). Transforming between Beck Depression Inventory and CORE-OM scores in routine clinical practice.
- Leichsenring, F., Kunst, H., & Hoyer, J. (2003). Borderline personality organization in violent offenders: correlations of identity diffusion and primitive defense mechanisms with antisocial features, neuroticism, and interpersonal problems. *Bull Menninger Clin*, 67(4), 314-327.

- Lewin, K. (1936). *Principles of topological psychology*. New York: McGraw-Hill.
- Lewis, G., & Appleby, L. (1988). Personality disorder: the patients psychiatrists dislike. *Br J Psychiatry*, 153(1), 44-49.
- Lewis, R. J. (1999). *Reliability and Validity: Meaning and measurement*. Paper presented at the Annual meeting of the American Society for Academic Emergency Medicine, Boston, Massachusetts.
- Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest*, 1(2), 27-66.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, California: Sage.
- Manfield, P. (1992). *Split self, split object: Understanding and treating borderline, narcissistic and schizoid disorders*. Northvale, NJ, USA: Jason Aronson.
- Martindale, C. (1975). *Romantic progression: The psychology of literary history*. Washington, DC: Hemisphere.
- Masling, J. M. (1997). On the nature and utility of projective tests and objective tests. *Journal of Personality Assessment*, 69(2), 257-270.
- Maxwell, J. A. (1996). *Qualitative research design*. London: Sage.
- McArdle, S., & Byrt, R. (2001). Fiction, poetry and mental health: expressive and therapeutic uses of literature. *J Psychiatr Ment Health Nurs*, 8(6), 517-524.
- McNiff, S. (1998). *Art-based research*. London: Jessica Kingsley.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). London: Sage.
- Millon, T. (1981). *Disorders of personality: DSM-III Axis-II*. New York: John Wiley and Sons.
- Millon, T., Davies, R., & Millon, C. (1997). *Millon Clinical Multi-axial Inventory-III manual* (2nd ed.). Minneapolis: National Computer Systems.
- Morgan, D. L. (1997). *Focus groups as qualitative research*. Thousand Oaks, California: Sage.
- Moriarty, N., Stough, C., Tidmarsh, P., Eger, D., & Dennison, S. (2001). Deficits in emotional intelligence underlying adolescent sex offending. *J Adolesc*, 24(6), 743-751.
- Murray, H. A. (1943). *Thematic Apperception Test manual*. Cambridge, Mass.: Harvard University.

- NIMHE. (2003). *Personality disorder: No longer a diagnosis of exclusion*. London: Department of Health.
- Ozarin, L. (2001). Moral insanity: A brief history. *Psychiatric News*, 36(10), 21-.
- Paley, G., Myers, J., Patrick, S., Reid, E., & Shapiro, D. A. (2003). Practice Development in psychological interventions: Mental health nurse involvement in the Conversational Model of psychotherapy. *Journal of Psychiatric and Mental Health Nursing*, 10, 494-498.
- Péladeau, N. (1998). WordStat: Content analysis module for SIMSTAT. Montreal: Provalis Research.
- Pendzik, S. (2003). Six keys for assessment in drama therapy. *The Arts in Psychotherapy*, 30(2), 91-99.
- Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, 77(6), 1296-1312.
- Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *BMJ*, 316(7139), 1236-1238.
- Prichard, J. C. (1835). A treatise on insanity. In M. H. Stone (Ed.), *In: Essential papers on borderline disorders (collection published 1986)* (pp. 14-16). New York: New York University Press.
- Propp, V. I. (1968). *Morphology of the folktale* (L. A. Wagner, Trans. 2nd revised ed.). Austin: University of Texas Press.
- Rennie, D. L. (1994). Storytelling in psychotherapy: The client's subjective experience. *Psychotherapy*, 31, 234-243.
- Richards, T. J., & Richards, L. (1991). The NUD*IST qualitative data analysis system. *Qualitative Sociology*, 14(4), 307-324.
- Rickers-Ovsiankina, M. (1992). Some theoretical considerations regarding the Rorschach method. In E. I. Megargee & C. D. Spielberger (Eds.), *Personality assessment in America: A retrospective on the fiftieth anniversary of the Society for Personality Assessment* (pp. 24-33). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Rogers, R. (2001). Structured Clinical Interview for DSM-IV personality disorders (SCID-II) and other Axis II interviews, *Chapter 9, Handbook of diagnostic and structured interviewing* (pp. 237-247). New York: Guilford Press.
- Rorschach, H. (1921). *Psychodiagnostik. Methodik und Ergebnisse eines wahrnehmungsdiagnostischen Experiments (Deutenlassen von Zufallsformen)*. Bern: Bircher.
- Rosenman, S., Korten, A., Medway, J., & Evans, M. (2003). Dimensional vs. categorical diagnosis in psychosis. *Acta Psychiatrica Scandinavica*, 107, 378-384.

- Rosse, I. C. (1890). Clinical evidences of borderland insanity. *Journal of Nervous and Mental Disease*, 17, 669-683.
- Rubin, E. (1915). *Synoplevde Figurer*. Copenhagen: Gyldendalske.
- Ryle, A. (1990). *Cognitive-Analytic Therapy: Active participation in change*. Chichester: John Wiley and Sons.
- Salmon, P. (2003). How do we recognise good research? *The Psychologist*, 16(1), 24-27.
- Sansone, R., Songer, D., & Gaither, G. (2001). Diagnostic approaches to borderline personality and their relationship to self-harm behavior. *International Journal of Psychiatry in Clinical Practice*, 5(4), 273-277.
- Scarpa, A., Luscher, K. A., Smalley, K. J., Pilkonis, P. A., Kim, Y., & Williams, W. C. (1999). Screening for personality disorders in a nonclinical population. *J Personal Disord*, 13(4), 345-360.
- Schoonjans, F. (2004). MedCalc (Version 7.5.0.0). www.medcalc.be.
- Shulman, D. G., McCarthy, E. C., & Ferguson, G. R. (1988). The projective assessment of narcissism: Development, reliability, and validity of the N-P. *Psychoanalytic Psychology*, 5(3), 285-297.
- Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.
- Singleton, N., Bumpstead, R., O'Brien, M., Lee, A., & Meltzer, H. (2003). Psychiatric morbidity among adults living in private households, 2000. *International Review of Psychiatry*, 15(1-2), 65-73.
- Snyder, C. R., & Shenkel, R. J. (1975). The P.T. Barnum effect. *Psychology Today*, 8(10), 52-54.
- Snyder, C. R., & Shenkel, R. J. (1976). Effects of 'favorability', modality and relevance on acceptance of general personality interpretations prior to and after receiving diagnostic feedback. *Journal of Consulting and Clinical Psychology*, 44, 34-41.
- Spradley, J. P. (1979). *The ethnographic interview*. Fort Worth, Texas: Harcourt Brace Jovanovich College Publishers.
- Stein, M. I. (1955). *The Thematic Apperception Test. An introductory manual for its clinical use with adults*. (Revised ed.). Cambridge, Mass.: Addison-Wesley.
- Stern, A. (1938). Psychoanalytic investigation of and therapy in the border line group of neuroses. *The Psychoanalytic Quarterly*, 7, 467-489.
- Stevenson, J., & Meares, R. (1999). Psychotherapy with borderline patients: II. A preliminary cost benefit study. *Aust N Z J Psychiatry*, 33(4), 473-477; discussion 478-481.

- Stigler, M., & Pokorny, D. (2001). Emotions and primary process in guided imagery psychotherapy: computerized text-analytic measures. *Psychotherapy Research, 11*(4), 415-431.
- Stirman, S. W., & Pennebaker, J. W. (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine, 63*(4), 517-522.
- Stone, M. H. (1986). The Borderline syndrome: Evolution of the term, genetic aspects and prognosis. In M. H. Stone (Ed.), *Essential papers on borderline disorders*. New York: New York University Press.
- Stone, P., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1966). *The General Inquirer: A computer approach to content analysis*: MIT Press.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing Grounded Theory* (2nd ed.). Thousand Oaks, California: Sage.
- Streiner, D. L., & Norman, G. R. (1995). *Health measurement scales: A practical guide to their development and use* (2nd ed.). Oxford: Oxford University Press.
- Szasz, T. (1991). Diagnoses are not diseases. *The Lancet, 338*, 1574-1576.
- Teasdale, J. D., & Barnard, P. J. (1995). *Affect, cognition, and change: Remodelling depressive thought*. Hove, East Sussex: Lawrence Erlbaum Associates.
- Tesnière, L. (1959). *Eléments de syntaxe structurale*. Paris: C. Klincksieck.
- Torgersen, S., Skre, I., Onstad, S., Edvardsen, J., & Kringlen, E. (1993). The psychometric-genetic structure of DSM-III-R personality disorder criteria. *Journal of Personality Disorder, 7*, 196-213.
- Urbaniak, G. C., & Plous, S. (2001). *Research randomiser*. Social Psychology Network. Retrieved 10 December, 2001, from the World Wide Web: <http://www.randomizer.org/form.htm>
- Vane, J. R. (1981). The Thematic Apperception Test: A review. *Clinical Psychology Review, 1*, 319-346.
- West, M., Pettem, O., Sheldon-Keller, A., Rose, M. S., & Cawthorpe, D. (1995). A story telling method for the assessment of attachment in adults. *British Journal of Projective Psychology, 41*(1), 30-48.
- Westen, D., Barends, A., Leigh, J., Mendel, M., & Silbert, D. (1990). *Social Cognition and Object Relations Scale (SCORS): Manual for coding interview data*. Ann Arbor: University of Michigan.
- Westen, D., Heim, A. K., Morrison, K., Patterson, M., & Campbell, L. (in press). Classifying and diagnosing psychopathology: A prototype matching approach. In L. Beutler & M. Malik (Eds.), *Rethinking the DSM: Psychological perspectives*. Washington, DC: American Psychological Association.

- Westen, D., Lohr, N. E., Silk, K. R., Gold, L., & Kerber, K. (1990). Object relations and social cognition in borderlines, major depressives and normals: A Thematic Apperception Test analysis. *Psychological Assessment*, 2, 355-364.
- Westen, D., Ludolph, P., Lerner, H., Ruffins, S., & Wiss, F. C. (1990). Object relations in borderline adolescents. *J Am Acad Child Adolesc Psychiatry*, 29(3), 338-348.
- Winnicott, D. W. (1971). *Playing and reality*. London: Tavistock.
- World Health Organisation. (1995). *International Personality Disorder Examination IPDE Manual, DSM-IV Module*. New York: Author.