

NOT FOR WIDER CIRCULATION OR REPRODUCTION

**Monitoring positive mental
health in Scotland: validating
the Affectometer 2 scale and
developing the Warwick-
Edinburgh Mental Well-Being
Scale for the UK**

September 2006

Ruth Tennant, University of Warwick
Ruth Fishwick, University of Warwick
Stephen Platt, University of Edinburgh
Stephen Joseph, University of Nottingham
Sarah Stewart-Brown, University of Warwick

Contents

Acknowledgements

Summary

1. Introduction	1
1.1 Background	1
1.2 The origins of the Affectometer.....	2
1.3 Scale validation processes	3
1.4 Initial validation of Affectometer 2 in a Scottish population.....	5
1.5 Research context.....	6
1.6 Research objectives	7
1.7 Research process.....	8
1.8 Ethical approval.....	8
2 Phase 1: Review of the literature relating to the validation and uses of Affectometer 1 & 2.....	10
2.1 Aims.....	10
2.2 Methodology.....	10
2.2.1 Methodology limitations	10
2.3 Results	11
2.3.1 Results of validation studies on Affectometer 2	11
2.3.2 How have Affectometer 1 and 2 been used?.....	11
2.3.3 Strengths and weaknesses of Affectometer 2.....	13
2.4 Discussion	14
2.4.1 Validation of Affectometer 2.....	14
2.4.2 Uses of Affectometer 2	14
3 Phase 2: Face validation of Affectometer 2 with focus groups.....	16
3.1 Aims.....	16
3.2 Methods.....	16
3.2.1 Recruitment	16
3.2.2 Focus group approach.....	17
3.2.3 Analysis	18
3.3 Results	18
3.3.1 What Affectometer 2 is measuring.....	18
3.3.2 Ease of completion	19
3.3.3 Definitions of mental health.....	20
3.3.3.1 <i>Mental health</i>	20
3.3.3.2 <i>Well-being</i>	21
3.3.3.3 <i>Mental illness and mental well-being</i>	21
3.3.4 Extra items.....	22
3.3.5 Sentences versus words.....	22
3.3.6 Differences between focus groups results	22
3.3.7 Focus group participants.....	23
3.4 Key conclusions.....	23
4 Phase 3: Psychometric testing of Affectometer 2 with student samples...25	25
4.1 Aims.....	25

NOT FOR WIDER CIRCULATION OR REPRODUCTION

4.2	Method	25
4.2.1	Ethics approval and consent.....	25
4.2.2	Sample size & questionnaire pack administration.....	25
4.2.3	Assessing construct validity	26
4.2.4	Relationship between positive and negative items in Affectometer 2	29
4.2.5	Assessing test-retest reliability.....	29
4.2.6	Assessing response bias	29
4.2.7	Data Entry and Quality.....	30
4.2.8	Statistical tests.....	30
4.3	Results	30
4.3.1	Data analysis	30
4.3.2	Response rates.....	30
4.3.3	Affectometer 2: descriptive data	31
4.3.4	Internal consistency	34
4.3.5	Construct validity	34
4.3.6	Relationship between positive and negative items of Affectometer 2	35
4.3.7	Test-retest reliability.....	35
4.3.8	Response bias.....	36
4.4	Discussion	37
4.4.1	Internal consistency	37
4.4.2	Construct validity	38
4.4.3	Test-retest reliability.....	38
4.4.4	Response bias.....	38
4.5	Key conclusions.....	38
5	Phase 4: Development of the Warwick-Edinburgh Mental Well-being Scale	39
5.1	Aims.....	39
5.2	Method	39
5.3	Results	39
5.3.1	Removing scale items.....	39
5.3.2	New scale development.....	40
5.4	Conclusion - Warwick-Edinburgh Mental Well-Being Scale.....	40
6	Phase 5: Psychometric testing of the Warwick-Edinburgh Mental Well-being Scale with student samples	42
6.1	Aims.....	42
6.2	Method	42
6.2.1	Questionnaire Packs.....	42
6.2.2	Construct validity	43
6.2.3	Analyses	43
6.3	Results	43
6.3.1	Data analysis	43
6.3.2	Response rates.....	43
6.3.3	Descriptive data	44
6.3.4	Factor analysis.....	47
6.3.5	Internal consistency	47
6.3.6	Construct validity	48
6.3.7	Test-retest reliability.....	49
6.3.8	Response bias.....	49

NOT FOR WIDER CIRCULATION OR REPRODUCTION

6.3.9	Correlation between Affectometer 2 and WEMWBS.....	50
6.4	Discussion of WEMWBS findings	50
6.4.1	Descriptive information	50
6.4.2	Internal consistency	51
6.4.3	Construct validity	51
6.4.4	Test-retest reliability.....	51
6.4.5	Response bias.....	51
6.4.6	Correlation between Affectometer 2 and WEMWBS.....	52
7	Face validity of WEMWBS	53
7.1	Aims.....	53
7.2	Method	53
7.3	Results	53
7.3.1	What the tool is measuring	53
7.3.2	Ease of completion	54
7.3.3	How the scale might be used.....	54
7.3.4	Limitations on scale use.....	54
7.3.5	Scale time-scale	55
7.3.6	Definitions of mental health.....	55
	7.3.6.1 <i>Mental health</i>	55
	7.3.6.2 <i>Mental health and physical health</i>	56
	7.3.6.3 <i>Mental illness and mental well-being</i>	56
7.4	Key Conclusions.....	56
8	General discussion of findings	58
8.1	Underlying purpose of Affectometer 2 and WEMWBS	58
8.2	Is the scale measuring 'positive mental health'?.....	58
8.3	Do Affectometer 2 and WEMWBS measure the same concept?.....	59
8.4	Scale completion and use	59
8.5	Strengths and weaknesses of this research	60
9	Overall conclusions	62
9.1	Suitability of WEMWBS as a measure of positive mental health at a population level	62
9.2	Potential future uses of WEMWBS	62
9.3	Suitability of WEMWBS for use at an individual level	63
9.4	Affectometer 2 or WEMWBS?	63
	References	64
	Appendices.....	67
	Appendix 1: Affectometer 2	68
	Appendix 2: Analysis of HEPS dataset.....	70
	Appendix 3: Focus group information sheet	91
	Appendix 4: Focus group consent form	93
	Appendix 5: Affectometer 2 - Schedule for Focus Groups.....	94
	Appendix 6: Affectometer 2 questionnaire information sheet.....	96
	Appendix 7: Student Consent form	98
	Appendix 8: The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS).....	99

Acknowledgements

This work was commissioned by Dr Jane Parkinson (Public Health Adviser, Mental Health Indicators, NHS Health Scotland) to support work for the Scottish Executive National Programme for Improving Mental Health and Well-being.

We would like to acknowledge the contribution and support of members of our Advisory Group:

- Prof Glynn Lewis, Department of Psychiatry, University of Bristol
- Dr Jane Parkinson, NHS Health Scotland
- Prof Jenny Secker, Institute of Health & Social Care, Anglia Ruskin University
- Prof Stephen Stansfeld, Queen Mary College, University of London
- Prof Scott Weich, Warwick Medical School, University of Warwick.

We are grateful to the focus group participants and students of the Universities of Warwick and Edinburgh who gave up their time to contribute to this research project, and to Alan Ross, of Ross Consulting who set up and ran the focus groups.

Finally we are grateful to Dr R Flett for permission to use Affectometer 2, and to him and his colleague Dr R Kammann for developing this scale which has proved to be ahead of its time.

Summary

The aim of this project was to test the face validity, content validity, construct validity (convergent and divergent) and reliability of the Affectometer 2 scale in the UK, to assess its suitability for measuring positive mental health (of adults) at a population level in Scotland and other parts of the UK, to modify the scale if appropriate, reducing its length and increasing its validity and reliability, and to validate the shortened version.

The starting point for the project was a literature review (Stewart-Brown, 2002) that highlighted the lack of scales available to measure positive mental health and identified one - Affectometer 2 (Kammann & Flett, 1983a) that had intuitive appeal and seemed to correspond to current definitions of positive mental health. Developed in New Zealand in 1983, this scale had undergone validation in that country at that time, but had since been used only in small scale projects, with very limited use in the UK. The scale consists of 40-items: 20 words and 20 sentences, including positively and negatively-phrased items.

The face validity of Affectometer 2 was tested by discussing the tool with focus groups from different socio-economic backgrounds in Scotland and England. The members of the groups were also asked to comment on the wording of individual items and report any problems they had experienced with understanding and responding to scale items.

Many members of the groups struggled with the concept of positive mental health and tended to equate 'mental health' with depression or mental illness. They had some minor comments on wording of individual items, but overall found the scale acceptable, interesting and easy to complete.

Construct validity, stability over time and internal consistency were assessed using data collected from students at the Universities of Warwick and Edinburgh. In these samples, Affectometer 2 showed good levels of construct validity, correlating well with scales measuring psychological well-being, life satisfaction and mental illness, largely in line with prior expectations. In this phase of testing, Affectometer 2 showed higher levels of response bias than other comparable scales and a very high level of internal consistency, the latter suggesting some redundancy of items in the scale.

Data from the above phases of the project were used, together with data gathered on Affectometer 2 from a Scottish population survey (the Health Education Population Survey 2002), as the basis for developing a new scale the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS). This new scale has fourteen positively worded items covering most aspects of positive mental health, including both hedonic (subjective experience of happiness and life satisfaction) and eudaimonic (psychological functioning, relationships with others) perspectives. The construct validity and reliability of WEMWBS was tested on student samples at the Universities of Warwick and Edinburgh and compared to a series of scales measuring different aspects of mental health and well-being. It showed moderate to high levels of construct validity with the other comparable scales and less response bias than Affectometer 2. The internal consistency of WEMWBS suggested less risk of item redundancy than had been observed with Affectometer 2. However, internal

NOT FOR WIDER CIRCULATION OR REPRODUCTION

consistency remained high enough to suggest that further reduction in the number of items might be possible. In the final phase of the research the revised scale was discussed with two additional focus groups to assess its face validity and acceptability.

Conclusions

Affectometer 2 performed reasonably well as a focus measure of positive mental health in the UK population. The main limiting factors were a tendency to response bias and its length. In contrast, WEMWBS showed significant improvements: as well as being shorter and easier to complete than Affectometer 2, it was less prone to response bias than Affectometer 2. In addition, its exclusively positive content appears to fit better with the focus group's views on what constitutes positive mental health. WEMWBS also retains some of the key strengths of Affectometer 2, appearing to have good face validity, and appropriate levels of convergent and divergent validity with a range of comparable scales.

This research suggests that WEMWBS is likely to be a user-friendly and psychometrically sound tool for monitoring positive mental health at a population level in Scotland and the UK. Research is currently underway to test the scale's sensitivity to change; this will indicate whether or not the scale can be used to test changes in positive mental health before and after a programme or intervention. Additional research is also planned to repeat the analyses carried out in this report on a large population sample in Scotland. This will clarify the way the scale varies with factors such as social class, age, marital status, income level and other known correlates of mental health, and contribute to the development of population norms.

1. Introduction

1.1 Background

Throughout the developed world, interest in the concept of positive mental health is growing among policy makers and health service providers. Some of this interest is driven by the public who are concerned that health services should look beyond the prevention and treatment of disease to embrace the development of well-being (DH 2005). Policy makers are beginning to act on such demands, developing mental health promotion programmes and initiatives (European Commission 2005, Scottish Executive 2003). Monitoring the impact of these policies requires validated scales which reflect current concepts of positive mental health.

Positive mental health is, however, a complex construct and there is still debate about its precise nature. The terminology used in discussion about this nature can, in some situations, reduce rather than enhance clarity and may fuel debate and disagreement. For example, both policy makers and academics continue to debate the extent to which happiness and life-satisfaction are components of positive mental health and these debates get caught up in disagreement as to whether the latter are legitimate goals of public services. Some disciplines – for example economists - have embraced such policy goals, others, particularly those involved in the health services, have not. To add to the confusion, the term ‘positive mental health’ is often used, in both policy and academic literature, interchangeably with the term mental well-being, and mental well-being may be represented simply as ‘well-being’.^α Some things are, however, largely agreed, perhaps most importantly that positive mental health encompasses more than the absence of mental illness. There is also increasing agreement that it covers both experience and functioning and that it has two distinct perspectives:- the hedonic perspective which focuses on the subjective experience of happiness and life satisfaction and the eudaimonic perspective favoured by positive psychologists, focusing on psychological functioning, good relationships with others and self realisation (Ryan and Deci 2001). These two perspectives have informed distinct bodies of research in positive mental health and their relative importance continues to be the subject of debate. The distinction between the two is, however, much less obvious in the literature relating to poor mental health where items measuring affect (subjective feelings of lack of well-being e.g. feeling sad/anxious) are usually combined with items measuring lack of psychological functioning (feeling useless, not facing problems, poor decision making and poor interpersonal relationships) (Goldberg and Williams 1988) suggesting that poor mental health is accepted as involving limitations in both eudaimonic and hedonic well-being. Within this context the World Health Organisation’s definition of positive mental health, ‘a positive state of well-being, one which allows individuals to fully engage with others, cope with the stresses of life and realise their abilities’ is a good starting point (World Health Organisation, 2001).

In the context of this lack of consensus, the development of a measure of positive mental health is challenging. In the context of the need to re-orient services away from disease to encompass a focus on well-being, such development is also

^α In this report the term positive mental health is used the majority of the time although in certain places mental well-being or well-being are used instead

NOT FOR WIDER CIRCULATION OR REPRODUCTION

necessary. This report describes the results of a project which aims to rise to this challenge by producing a measure for adults which is good enough for purpose at the beginning of the twenty first century. In this way it needs to reflect the academic literature on the subject. It also needs to be acceptable to the general public and to be in accord with their current understanding of the concept, whilst accepting that the general public does not speak with one voice nor think with one head. As such the measure which was developed does not aim to be the last word on the subject, it aims only to fill a gap by supporting current evaluation of policies, programmes and initiatives to promote positive mental health.

1.2 The origins of the Affectometer

Historically, mental health measurement scales have focused on psychiatric morbidity, dividing the population into those who meet the criteria for diagnosis of mental illness and those who do not. These scales are not well-suited to measuring positive mental health at a population level as they show 'ceiling effects' with most people scoring the maximum possible score. They are therefore unable to distinguish average from good mental health (Stewart-Brown, 2002).

A recent review of scales of positive mental health (Stewart-Brown 2002) identified one – Affectometer 2 (Kammann & Flett, 1983a) – with intuitive appeal that appears to relate well to current definitions of positive mental health (e.g. WHO 2001) and covers both subjective well-being and psychological functioning. This scale, developed in New Zealand in 1983, covers both hedonic and eudaimonic aspects of positive mental health including: confidence, usefulness, interest in life, problem solving, autonomy, positive relationships, thinking clearly and creatively, as well as energy, happiness and optimism (Table 1).

The scale was developed through the scrutiny of 435 adjectives and sentences derived from a review of the literature existing at that time (relating to positive mental health). These items were reduced to 96 items to create Affectometer 1 (Robinson *et al*, 1991). Affectometer 1 was tested in a general population sample and reduced to the 40-item Affectometer 2 which comprised of 20 positive items and 20 negative items each split into 10 sentences and 10 adjectives.

Respondents complete each item using a 5-point Likert scale indicating how much they agree with each word or statement, ranging from 0 ('not at all') to 4 ('all of the time'). Overall scores are calculated by subtracting the sum of all negative items from the sum of all positive items, giving a possible score range of between +80 and -80.

The scoring system for Affectometer 2 reflects one of the instrument's underlying theoretical principles that an individual's mental health status is determined by the degree to which positive feelings, attitudes and beliefs outweigh negative ones. A copy of the scale can be found in Appendix 1.

The scale is free for use in non-commercial organisations although authorisation to use the scale should be sought from the scale's author (see Appendix 1 for contact details). The scale has been used in the UK as part of the Health Education Board for Scotland's (now part of NHS Health Scotland) Health Education Population Survey in

NOT FOR WIDER CIRCULATION OR REPRODUCTION

2002 and 2005. Minor modifications were made to the scale’s wording to make it more appropriate for use with a UK population.

Table 1: Affectometer 2 items*

Positive sentences	Positive adjectives	Negative sentences	Negative adjectives
My life is on the right track	Satisfied	I wish I could change some part of my life	Discontented
My future looks good	Optimistic	I feel as though the best years of my life are over	Hopeless
I like myself	Useful	I feel there must be something wrong with me	Insignificant
I can handle any problems that come up	Confident	I feel like a failure	Helpless
I feel loved and trusted	Understood	I have been left alone when I don’t want to be	Lonely
I feel close to people around me	Interested in other people	I have lost interest in other people and don’t care about them	Withdrawn
I feel I can do whatever I want to do	Relaxed about things	My life seems stuck in a rut	Tense
I have energy to spare	Enthusiastic	I can’t be bothered doing anything	Depressed
I smile and laugh a lot	Good natured	Nothing seems very much fun anymore	Impatient
I have been thinking clearly and creatively	Clear headed	My thoughts have been going round in useless circles	Confused

* Note UK wording is shown here

Since the 2002 review of positive mental health scales was published (Stewart-Brown 2002), a more detailed review of instruments aiming to measure aspects of positive mental health (Speight et al, forthcoming) has confirmed Affectometer 2’s potential use for monitoring purposes in Scotland, but also highlighted the need for the further development of validated measures to assess positive mental health.

1.3 Scale validation processes

Before a scale is adopted, it needs to undergo a process of validation. This is a structured process to test whether the scale is robust and whether it is suitable for wider use. The key processes in scale validation are summarised in Table 2 (Lohr et al, 1996).

Table 2: Key processes in scale validation*

Attribute	Meaning	Assessment method/ statistical process
1. Conceptual model: What is the scale purporting to measure?		
	a) Is there a clear rationale for the scale? b) Does it measure one concept which can be summarised in a single score? c) Are its measurement properties clearly described?	a) Description of the underlying purpose of the scale, drawing on relevant literature. b) Factor analysis, using principal component analysis or other relevant method. c) Description of how the scale is scored, descriptive statistics on how the scale performs (mean, standard deviation etc).
2. Reliability (internal consistency & test-retest reliability): The degree to which the scale is free from random error.		
2a) Internal consistency	Are different items in the scale measuring a consistent concept?	Cronbach's Alpha
2b) Test-retest reliability	Does the scale show appropriate levels of stability over time?	Completion of the scale by the same people at different points in time.
3. Validity: Does the scale measure what it says it does?		
3a) Face (or content) validity	Are the items in the scale suitable for the overall concept being measured?	Lay or expert-panels/ focus groups.
3b) Construct (convergent /divergent) validity	Are there logical relationships between the scale and other scales or factors known to affect the concept being measured (such as age or sex)	Correlations between scale under review and other scales measuring similar concepts (convergent validity) or different concepts (divergent validity). Statistically significant differences in scales scores between different groups.

Attribute	Meaning	Assessment method/ statistical process
3c) Criterion validity	How does the scale compare to a 'gold standard' measure?	Correlations between scale under review and 'gold standard'.
4. Susceptibility to bias: Do responses to the scale vary depending on the certain individual attributes?		
	Is the scale prone to response bias or to people tailoring their responses to present themselves in a more (or less) favourable way?	Correlation between the scale and scales that are designed to measure response bias.
5. Responsiveness to change: Do responses change as the level of health changes?		
	Is the scale capable of detecting change over time (for example, before and after an intervention)?	Change in scale scores before and after an event or intervention that is likely to affect the concept being measured.

* adapted from Lohr et al, 1996

Validation processes do not necessarily follow the order set out in this table; for example, face validity testing is usually carried out before a scale is used on a wider sample, although it may also be used to modify an existing scale. Not all of the steps outlined in this table will be relevant to all scales: for example, criterion validity is usually applied to scales such as depression scales, where the scale's performance can be compared to a recognised 'gold standard', such as a clinical diagnosis of depression. The lack of a 'gold standard' measure for concepts such as positive mental health and mental well-being means that this test is not usually applied to scales measuring these concepts. Responsiveness to change may also be tested only once the other validation tests have been completed.

1.4 Initial validation of Affectometer 2 in a Scottish population

Initial steps in validating Affectometer 2 were carried out before this research project was commissioned, using data collected in the fourth wave (2002) of the Scottish Health Education Population Survey (HEPS). For full details see Appendix 2. The latter study examined Affectometer 2's internal consistency using Cronbach's alpha and the underlying structure of the scale using principal component analysis (a form of factor analysis). Prior hypotheses about the expected association between scores on Affectometer 2 and the General Health Questionnaire 12 (GHQ-12) (also collected in the HEPS survey) and associations with socio-economic factors known to predict

NOT FOR WIDER CIRCULATION OR REPRODUCTION

poor mental health were developed to test the tool's construct validity. The performance of four different short versions of the scale (all 20 positive items; all 20 negative items; all 20 adjectives and all 20 statements) was also tested in the analyses. Eighty nine percent of respondents to this survey completed Affectometer 2 suggesting a high level of public acceptability. The scale showed high internal consistency (0.944), pointing to possible item redundancy. With the exception of gender, where no significant differences were found, correlations between the Affectometer 2 score and key socio-demographic variables were consistent with the findings of other population mental health surveys such as the UK longitudinal study of the mental health of adults living in private households in Great Britain (Singleton, 2003) (see Appendix 2).

The scale did not show a ceiling effect in this general population sample, indicating that the measure has the potential for documenting overall improvements in positive mental health at a population level. The principal component analysis was consistent with, but not unequivocally supportive of, a single underlying construct such as is necessary to justify the scoring method (subtracting the sum of negative items from positive items which assumes equal weighting of all states). Appendix 2 provides a more detailed account of the principal components analysis.

Whilst the results of these analyses were promising, it was not possible from the data to test the scale's test-retest reliability nor to compare the scale's performance to scales other than the GHQ-12. On the back of this initial research, NHS Health Scotland commissioned additional research (described in the rest of this report) to meet these gaps.

1.5 Research context

This research was commissioned by NHS Health Scotland as part of its mental health indicators programme (for information on the indicators programme see <http://www.healthscotland.com/understanding/population/mental-health-indicators.aspx>). This programme, which started in December 2003, aims to develop a core set of sustainable national mental health indicators for Scotland, in the first instance a set for adults. It is one of the "support activities" for the Scottish Executive's National Programme to Improve Mental Health and Well-being, as outlined in the National Programme's Action Plan 2003-6 (Scottish Executive 2003).

The main objectives to the mental health indicators project are to:

1. Determine a desirable set of defined mental health indicators;
2. Review relevant Scottish data currently collected nationally, to maximise the use of data already gathered;
3. Identify, and establish a consensus on, a set of practical indicators for which data are currently available;
4. Recommend new data that should be collected to fill gaps between the identified set of practical indicators and the desirable set;
5. Explore approaches to collect additional data to cover the identified gaps and work to develop the data collection systems for desired indicators where they do

not currently exist;

6. Ensure the sustainable collection of both current and new data.

Affectometer 2 was identified by the Advisory Group of the mental health indicators programme as having the potential to provide data that is important for NHS Health Scotland's mental health indicator set, but is currently lacking. As this scale had not been fully validated in the UK, this research was commissioned with the aim of validating the scale and enabling conclusions to be drawn about its suitability for use as a measure of population positive mental health in Scotland, and UK.

1.6 Research objectives

The objectives of this research were to:

- review the literature available on Affectometer 2, specifically including its advantages and disadvantages over other similar scales
- establish the face validity of Affectometer 2
- assess the performance of Affectometer 2 against other validated mental health and well-being measurement scales (convergent validity)
- assess whether Affectometer 2 is a suitable and appropriate scale to capture the missing data required for the mental health indicators set
- develop and validate a shorter version of Affectometer 2 if appropriate.

This research was conducted in six phases and the report covers each in turn setting out the methodology and results and discussing the findings.

The report on the first phase sets out what was known about Affectometer 2 and establishes the need for additional testing of the scale. It describes a literature review which outlines how Affectometer 2 has been used to date, and gives a brief description of its advantages and disadvantages. A fuller review of different positive mental health scales, also commissioned by NHS Health Scotland, should be read in parallel with this research (Speight et al. forthcoming).

The report on the second phase describes the process of testing the face validity of Affectometer 2 with focus groups in Scotland and England. As the scale's face validity had not been tested with potential users, this component was seen as an important way of establishing whether the scale was compatible with the public's perception of positive mental health, and to test the extent to which the scale is user-friendly.

The report on the third phase describes the process of carrying out psychometric testing on Affectometer 2 on student samples at Warwick and Edinburgh Universities. This process was used to assess the tool's construct validity by comparing its performance alongside similar scales, its internal consistency (whether individual items in the scale measure a similar construct), its test-retest reliability (whether the scale is stable when repeated on the same people at different points in time) and whether the scale is susceptible to response bias (people amending their response to

NOT FOR WIDER CIRCULATION OR REPRODUCTION

present themselves in a more – or less - favourable light). All these processes are standard methods of validating scales such as Affectometer 2.

The report on the fourth phase describes the process of developing a new scale (the Warwick-Edinburgh Mental Well-Being Scale) (WEMWBS), using the findings of the focus groups and the results of the psychometric testing outlined in phase 3.

The report on the fifth phase describes the process of carrying out psychometric testing on WEMWBS, and assessing its correlation with Affectometer 2. The comparative performance of Affectometer 2 and WEMWBS is discussed.

The report on the sixth phase describes two additional focus groups in Scotland and England which were carried out to assess the face validity of WEMWBS.

The final two sections of the report draw conclusions about the performance of the two scales, - the Affectometer 2 and WEMWBS – describing their suitability for use as measures of positive mental health for adults at both an individual and population level and highlighting some remaining gaps in the research.

1.7 Research process

This research was commissioned in April 2005 and was completed in June 2006. The research was carried out by a research team based at Warwick and Edinburgh Universities and was overseen by an Advisory Group (see acknowledgments) convened by NHS Health Scotland. The Advisory Group included experts in mental health (psychiatry, mental health promotion and positive psychology) and the main research techniques used in the project (scale validation and qualitative research), the commissioner and the project team. The Advisory Group met three times and also provided feedback at all stages of the research via email.

1.8 Ethical approval

Although this research did not require approval from an NHS Research Ethics Committee as it did not involve NHS patients, staff or facilities, it was submitted for approval to Warwick Medical School Ethics Committee, using standard NHS ethics committee procedures. The Committee placed a number of limitations on the study, requiring first that an explicit reference to 'mental health' be made in the information sheets in place of the term 'emotional well-being'. Focus group participants were given these information sheets before deciding whether or not to take part in the research. As the focus groups subsequently showed, in line with previous research, respondents generally equated the term 'mental health' with 'mental illness'. The information sheet may therefore have communicated a misleading impression about the true nature of the research and potentially influenced participants' responses. References in the study information sheet to helpline numbers for the Samaritans, which were also added at the request of the Ethics Committee, may have reinforced this impression.

The Ethics Committee initially queried the use of payments to members of the public taking part in the research, reflecting tensions between traditional interventional research where financial inducements to take part are considered unethical, and

NOT FOR WIDER CIRCULATION OR REPRODUCTION

research involving users, where payment to participants, reflecting the skills and knowledge that they bring to research, is standard practice. The Committee accepted the research team's position that making payments to participants to cover expenses was in line with current NHS guidance on patient and public involvement in research.

2 Phase 1: Review of the literature relating to the validation and uses of Affectometer 1 & 2

2.1 Aims

The first phase of this research involved a literature review, carried out in April 2005. The aims of the review were to:

- identify the ways that Affectometer 1 & 2 have been used since their development;
- identify any additional validation studies that have been carried out on Affectometer 2 since its original validation in New Zealand;
- identify any instances of use as a population scale of positive mental health;
- outline the advantages and disadvantages of Affectometer 2, drawing on the results of the literature review and the earlier review which assessed the content validity of Affectometer 2 compared with other scales of well-being (Stewart-Brown, 2002).

2.2 Methodology

A search was run on four electronic bibliographic databases (Psychinfo, CINAHL, EMBASE and MEDLINE), using the search-term 'Affectometer' in the title or abstract. These search engines were selected to cover the main disciplines (psychology, health and social care, nursing and allied health literature, and medicine) in which Affectometer 1 or 2 could have been used. No language or time restrictions were applied. All study types were included in the search. In addition, the scale's authors were contacted for information (published or unpublished) about Affectometer 2's original validation.

Articles were retrieved and reviewed individually. The criteria for including articles in the literature review was that the Affectometer (either version 1 or version 2) should have been used to measure any dimension of mental health (positive mental health, well-being, positive or negative mood) at an individual, group or population level, either at a single point in time or as a repeated measure. Dissertations identified by the four databases were not retrieved for cost reasons but dissertation abstracts were reviewed.

In the remainder of this section, we use the terminology adopted by the authors' of each study.

2.2.1 Methodology limitations

The search was restricted to electronic databases. No hand searches were carried out and experts in the field were not contacted because of time constraints. The search strategy was also limited to articles that included the term 'Affectometer' in either the title or abstract so may not have captured all the studies that used Affectometer 1 or 2 as one of a series of measures. However, several of the studies that were identified did use Affectometer 1 or 2 as just one of a series of measures and the search strategy used is likely to have identified any studies in which Affectometer 1 or 2 was the main focus of the research.

2.3 Results

Fifteen studies were identified, including the original validation study article written by Affectometer 2's authors (Table 3). Of these, ten were published in journals, four were unpublished dissertations and one was an unpublished handbook, written by the scale's authors. The majority of the studies had been carried out in Australia or New Zealand, suggesting limited diffusion beyond the scale's region of origin.

Table 3: Summary of search results

Database	No. of relevant returns	No. of unique returns
Medline	4	4
Psychinfo	7	6
EMBASE	5	0
CINAHL	7	4
<i>Other*</i>	-	1
TOTAL	-	15

** references identified by one of the authors of Affectometer 2*

2.3.1 Results of validation studies on Affectometer 2

Affectometer 2 was originally validated in New Zealand using a sample of 112 adults (Kammann & Flett, 1983a). This validation showed that the scale had an internal consistency (Cronbach's alpha) of 0.952, and a test-retest reliability of 0.83. Construct validity and response bias were only reported for the 96-item Affectometer 1 showing moderate to high correlations (0.6 to 0.7) against life satisfaction scales and general well-being scales, and high negative correlations with the Beck Depression Inventory.

2.3.2 How have Affectometer 1 and 2 been used?

One group of studies used Affectometer 1 or 2 to measure changes in some aspect of mental health before and after an event or intervention (Dennerstein, 2002; Lo, 1997; Peters, 2002; Raeburn, 1994; & Stein, 1996). This included assessing the impact of eye movement desensitisation and reprocessing, occupational therapy, and a life-style change programme on well-being, as well as changes in well-being associated with menopausal status (Table 4). Collectively these studies demonstrated that the instrument is sensitive to change in mental health states.

A second category of studies used Affectometer 1 or 2 to identify associations between a particular mood or state and mental health status, including associations between paternal responsibility and well-being, and between personality traits and well-being (Hossack, 1997; Kimweli, 2002; O'Byrne, 1997; Rand, 1982) (Table 4). These studies generally found associations in the directions expected.

No examples of Affectometer 1 or 2 being used at a general population level were found, although it has been used with a range of ages, socio-economic groups and people of different health status. No additional validation studies other than the original study carried out by Kammann and Flett were identified and no examples of

NOT FOR WIDER CIRCULATION OR REPRODUCTION

research carried out to explore the scale's face validity with its potential end-users were found.

The largest proportion of retrieved studies (n=8) were carried out in Australia or New Zealand. Four studies were carried out in North America (one in Canada, three in the U.S.). The remaining studies were carried out in South Africa (n=1) and Taiwan (n=1) (Table 4). We found no published examples of Affectometer 1 or 2 being used in the UK or the rest of Europe.

Table 4: Summary of uses of Affectometer 1 or 2

Study	Author* (country)	Use of Affectometer
Psychological well-being, mid-life and the menopause.	Dennerstein, 1994 (AU)	Affectometer 2 used to assess changes in positive and negative mood and overall well-being during menopausal transition
Factors contributing to positive mood during the menopausal transition.	Dennerstein, 2001 (AU)	As above
Effects of the menopausal transition and biopsychosocial factors on well-being.	Dennerstein, 2002a (AU)	As above
Hormones, moods, sexuality and the menopause transition.	Dennerstein, 2002b (AU)	As above
Psychosocial aspects of chronic lower leg ulceration in the elderly.	Flett, 1994 (NZ)	A 10-item version of Affectometer 2 used to evaluate psychological well-being in elderly people with leg-ulcers compared with a control group.
Dissertation on attachment, personality, sense of coherence and well-being in undergraduates.	Hossack, 1997 (Canada)	Affectometer 2, one of 12 scales used to assess the influence of personality traits on subjective well-being
Affectometer 2: a scale to measure current level of general happiness.	Kammann, 1983a (NZ)	Validation study of Affectometer 2
Sourcebook for measuring well-being with Affectometer 2.	Kammann, 1983b (NZ)	Handbook for using Affectometer 2. Includes data on validation of Affectometer 1 and 2.

NOT FOR WIDER CIRCULATION OR REPRODUCTION

Study	Author* (country)	Use of Affectometer
Dissertation on the relationship of gender-role development, college level and subjective well-being among Appalachians.	Kimweli, 2002 (US)	Affectometer 2, one of 3 scales used to measure well-being
Pilot study of the relationship between affective experiences during daily occupations and subjective well-being measures.	Lo, 1997 (Taiwan)	Affectometer 2 used to measure the levels of well-being of people receiving occupational therapy
Dissertation on paternal responsibility, involvement with children, and sense of well-being.	O'Byrne, 1997 (US)	Affectometer 2 used to assess paternal well-being (with Satisfaction with Life scale)
Implementation of EMDR with cancer patients.	Peters, 2002 (South Africa)	Affectometer 2 used to measure the impact of eye movement desensitisation and reprocessing on patients' positive-negative affect-balance
Psychological health and factors the courts seek to control in lesbian mother custody trials.	Rand, 1982 (US)	Affectometer 1 used to assess psychological health of participants
Superhealth Basic: development and evaluation of a low-cost community-based lifestyle change programme.	Raeburn, 1994 (NZ)	Programme evaluation: Affectometer 2 used to measure well-being before and after programme implementation
Dissertation on inter-institutional relocation of aged people.	Stein, 1996 (AU)	Affectometer 2 used to assess changes in psychological well-being before and after relocation of nursing home residents

* note only first author on paper is given

2.3.3 Strengths and weaknesses of Affectometer 2

This literature review and parallel research (see Stewart-Brown, 2002) identify a number of strengths and weaknesses of Affectometer 2 summarised in Table 5.

Table 5: Strengths and weaknesses of Affectometer

Strengths	Weaknesses
Used in a variety of settings to evaluate the impact of interventions/events on well-being and it is sensitive to change in mental health status.	Not widely used and no published instances of it being used in the UK or Europe (apart from in the Health Education Population Survey in Scotland, publication in progress Tennant <i>et al</i>).
Potentially flexible and adaptable – potential to develop a shorted version.	No published uses as a scale to measure positive mental health in a general population (apart from in the Health Education Population Survey in Scotland, publication in progress Tennant <i>et al</i>).
Used with participants from different socio-economic backgrounds, and with different health states therefore likely to be appropriate for use in a general population.	Not been re-validated since the original validation study which did not fully evaluate Affectometer 2 .
Captures a wide spectrum of aspects of positive mental health.	Longer than other comparable scales.

2.4 Discussion

2.4.1 Validation of Affectometer 2

The literature review found two sources of information on the validation of Affectometer 2, both of which contained the same material and were written by the scale’s authors. They showed that although information is available about the scale’s reliability (test-retest reliability and internal consistency, both of which were demonstrated to be good), there are some important gaps in knowledge about the scale’s performance. Although the authors provide favourable information on the 96-item Affectometer 1’s construct validity, this may not be transferable to the shorter 40-item Affectometer 2. There was also no evidence in the literature identified that the scale’s face validity has been tested with either a lay-panel or an expert-panel. This suggested that Affectometer 2 needed to be subject to a more robust validation, particularly of its face and construct validity, before it was adopted for wider use. Best practice in scale validation also suggests that full scale validation processes may be required before scales are used in different countries to the one in which they were developed (Lohr et al, 1996).

2.4.2 Uses of Affectometer 2

The literature review suggested that Affectometer 2 has not been widely used, particularly within Europe. This may reflect the fact that the original validation study was published in an Australian journal. It has, however, been successfully used to measure changes in different components of mental health status over time or before/after an event or intervention. It has also been used to measure associations between mental health status and different psychological states or traits. This review suggested that Affectometer 2 has not been used in population surveys (apart from

NOT FOR WIDER CIRCULATION OR REPRODUCTION

inclusion in HEPS, information on which has not yet been published) and that there remained a need to demonstrate its applicability in different contexts at a population level.

The literature review also demonstrated that the scale had not been validated since the original validation reported by Kammann & Flett (Kammann & Flett, 1983a). There were therefore substantial gaps in knowledge about how the scale performs, particularly how the scale compares to other similar scales, its face validity with end users in the twenty first century and whether it is possible to produce a shorter, validated version.

3 Phase 2: Face validation of Affectometer 2 with focus groups

3.1 Aims

The focus groups had six aims:

- to test what participants thought Affectometer 2 was designed to measure
- to test how easy participants found it to complete Affectometer 2 and in particular, whether they found the statement or word components of the scale easier
- to discuss participants' understanding of the terms 'mental health', 'mental well-being' and 'positive mental health'
- to discuss whether people thought Affectometer 2 was a measure of positive mental health
- to identify any items that people associated with mental well-being which were not covered by the scale
- to identify items that people did not understand, or did not think were relevant.

3.2 Methods

3.2.1 Recruitment

Concepts of mental health differ between different socio-economic and demographic groups. The focus groups therefore needed to test whether Affectometer 2 is a valid measure of positive mental health across the population. In order to assess this, nine groups (with a maximum of eight people in each) were selected using purposive sampling for focused discussion. These groups were selected to cover some of the attributes, such as socio-economic status, age and sex which are known to be associated with differences in mental health status (Singleton, 2003), plus one mental health user group.

Ethnicity and race are known to affect the way people conceptualise mental health and well-being (U.S. Department of Health and Human Services, 2001): after consultation with experts in the field of ethnicity and mental health, we concluded that to do justice to this subject, it would be necessary to investigate a wide range of linguistic and conceptual issues relating to mental health and ethnicity and that this was beyond the scope of this research. It was felt that a small number of focus groups with ethnic minority participants would be unlikely to capture the complex range of responses across different ethnic minority groups sufficiently. The research team therefore recommended to NHS Health Scotland that this topic should be the subject of a further research project.

The focus groups aimed to be diverse, fully inclusive and to include adults with a range of ages and socio-economic backgrounds (see Table 6). Community group organisers were contacted by phone by the research team using a common script. They were given verbal and written information about the purpose of the research and a copy of the project information sheet. They were asked to identify potential participants and to give them a copy of the project information sheet (Appendix 3). Focus groups took place between July and October 2005.

Table 6: Focus group details

Group	Location	Number of attendees
General women's group (25-64)	Leamington Spa, England	4
Young Men's Group (18-24)	Wolverhampton, England	3
Older Women's Group (65 +)	Leamington Spa, England	7
Older Men's Group (65 +)	Edinburgh, Scotland	8
General Men's group (65 +)	Edinburgh, Scotland	6
Young Women's Group (18-24)	Edinburgh, Scotland	8
Socio-economic groups C2, D, E Women	Musselburgh, Scotland	8
Socio-economic groups C2, D, E Men	Musselburgh, Scotland	7
Mental Health Service Users	Musselburgh, Scotland	5

3.2.2 Focus group approach

The focus group schedule was piloted twice. This included participants completing Affectometer 2, scoring their responses and creating an overall summary which could range from +80 to -80. Participants indicated that a negative score on Affectometer 2 could potentially upset people and as a result the scoring system was modified to remove this possibility for the remainder of the focus groups. No other changes were deemed necessary as a result of this pilot.

An independent consultant with experience in running focus groups was recruited to moderate all groups. A researcher took notes during the groups. All discussions were taped and transcribed by researchers at the University of Warwick.

Focus group discussions all took place in community venues and lasted a maximum of one and a half hours. Before the start of the discussion, participants were given a second verbal briefing about the project and asked whether they had read and understood the information sheet. They were then asked to sign two copies of the consent form (Appendix 4), one of which they kept and one was kept by the project researcher. Focus group participants were given £15 each to cover travel and any other expenses at the beginning of the session.

At the start of the group meeting, participants were asked to complete Affectometer 2 individually. Participants who experienced difficulty filling in Affectometer 2 were helped or given clarification by the researcher. After this, the moderator used a standard interview schedule (Appendix 5) to take the group through a series of questions that covered their initial (unprompted) responses to the scale and how easy they found it to complete. Participants were then asked to say what they thought the scale was attempting to measure. If necessary, participants were given prompts such as 'mental health' and 'mental well-being' and were asked to describe what they thought these terms meant and whether they thought Affectometer 2 was attempting to measure these concepts. Participants were also asked to identify any words or

NOT FOR WIDER CIRCULATION OR REPRODUCTION

phrases that they associated with good mental health that were not covered by Affectometer 2. At the end of the session, participants were asked to calculate their score and were shown a chart of the distribution of scores derived from analysis of the HEPS dataset (see Appendix 2).

3.2.3 Analysis

The interview schedule used in the focus groups was modified to develop a coding frame to which additional items were added by two researchers. Two transcripts were independently coded and compared by the same researchers. No significant changes to the coding frame were required at this point. All transcripts were coded using qualitative software NVIVO 2.

Summaries of the key points raised by the groups were sent out for comment and corroboration to all members of the groups that took part in this phase of research. No comments were received back from any members of the groups.

3.3 Results

3.3.1 What Affectometer 2 is measuring

Participants were asked to express their initial responses to Affectometer 2 and to describe what they thought it was attempting to measure.

There was a general consensus across the groups that the scale was 'thought-provoking' and that it tapped into feelings that they did not generally think about. Participants used words such as 'heavy' and 'deep' to describe their first thoughts about the scale, although the process of completing it did not appear to be threatening or difficult to any of the respondents. Two of the women's groups (women aged 18-24 and socio-economic group C2, D and E) appeared to find the process of completing the scale positive, saying that it had required them to think about things that people did not usually ask them about:

"Ken (you know), naeone (no-one) says to you how are you feeling, do you feel good-natured, lonely or whatever. So, you're no really asked that everyday. So, ken, you look at it and how it's, you think about it. 'Cause you're never really asked that everyday, so you have to think about (about) that."

(Women 18-24)

Although the scale contains an equal number of positive and negative items, many respondents, particularly in the young women's and the women's socio-economic groups C2, D and E, thought that the scale was measuring depression:

"It's mainly asking you whether you felt doon (down)."

(Women 18-24)

There appeared to be a consensus across the groups that Affectometer 2 is attempting to measure a concept, which relates to mental health although participants used a wide range of expressions to describe this concept:

NOT FOR WIDER CIRCULATION OR REPRODUCTION

“I think what it's trying to do is to try and find out where we are at this present moment by looking in the past, but I'd rather look to the future.”

(Men socio-economic groups C2, D, E)

Several respondents suggested that it could be used as a population tool to identify differences in mental health status between different groups and different areas:

“I would say it's more to find out ways of improving people's mental health as opposed to finding out say, about the nation's health. You know, to find out how the NHS could improve services and...”

(Men socio-economic groups C2, D, E)

“Some of them you've got 20's, 25-60 and you've got 65 and over. You've got male and female as well, so I my assumption is it's trying to check out the mentality between... or what kind of mental state these certain groups are in.”

(Men 18-24)

“When you were saying that... talking about that, it occurred to me I wonder if it is some kind of survey to assess different areas of the community within the country, so you would go to a deprived area and you would assess all of these people to find out possibly where they're coming from and to try and work out why they do certain things.”

(Women 25-64)

Overall, however, focus groups found it difficult to describe accurately what Affectometer 2 is measuring. This appears to reflect the prevailing view that 'mental health' is synonymous with mental illness and the lack of widely-recognised descriptors of positive mental states: concepts such as 'emotional health', 'well-being' or 'positive mental health' were generally not familiar to participants. However, participants did appear to recognise that the scale was attempting to measure a concept relating to their emotional state, described by one participant as 'how you feel in yourself'.

3.3.2 *Ease of completion*

There was agreement across all groups that the scale was straightforward and easy to complete and the length of the scale did not appear to be a problem in this context. However, both groups of older people, and the mental health service users questioned their own ability to complete the scale, particularly the older men, who were concerned about their ability to recall the previous two weeks:

“I can think back... a week back and when I go back the last two weeks it's sort of 'What did I do?'”

(Men over 65)

The two lower socio-economic groups and the women's group aged 18-24 felt that Affectometer 2 was quite depressing to complete and came across as negative. This may be because the final item of the scale is a negative item ('insignificant'). Several participants felt that this was a particularly difficult concept and, because it was the last item to be completed, could have a negative effect on respondents' mood:

NOT FOR WIDER CIRCULATION OR REPRODUCTION

“I don't mind being asked it but I think it's cruel tae (to), and I'm trying to put an answer to that and then somebody might think I'm insignificant that, I don't know but I think it's a very difficult question to answer.”

(Women socio-economic groups C2, D, E)

“Maybe 'cause it's the last word, insignificant, ken (you know), you're left with that, the last word.”

(Women socio-economic groups C2, D, E)

Both older groups struggled to follow an item along the correct line in order to answer it:

“Participant: I altered two but I just had a senior moment and filled in the wrong place and then I checked...”

Interviewer: Yes... OK.

Participant: ...Yes I did that and crossed them out again.”

(Women over 65)

One male from the lower socio-economic men's group found it difficult to differentiate between 'occasionally' and 'some of the time' in the Likert response scale:

“I had a bit of difficulty actually in between 'occasionally' and 'some of the time', because occasionally and some of the time strikes me as very similar. You know, if it's some of the time or occasionally well, some of the time could be occasionally so I just felt that those two things were confusing.”

(Men socio-economic groups C2, D, E)

3.3.3 *Definitions of mental health*

Following a discussion about what participants thought the scale was attempting to measure, each group was asked how they defined 'mental health'. The focus group moderator tested responses to different words that are associated with mental state, including 'mental health', 'mental illness', 'well-being', 'emotional health' and 'positive mental health.'

This section of the discussion generated a wide range of ideas and comments and perhaps unsurprisingly, there was no clear consensus about what each of these terms meant.

3.3.3.1 *Mental health*

'Mental health' was viewed by the majority of the focus groups as an umbrella term for a range of mental illnesses such as depression, schizophrenia and manic depression. Many groups talked about the role of medication in controlling mental health problems. Some participants described mental health as being controlled by the 'brain' but recognised a parallel concept of 'emotional health', which related to what they described as 'feelings.'

Women and men from lower socio-economic groups recognised that mental health can be affected by life experiences and circumstances. Women from lower socio-

NOT FOR WIDER CIRCULATION OR REPRODUCTION

economic groups explained that those who had a high Affectometer 2 score would display good mental health. Young women (18-24) viewed depression as a mental illness, a state of mind and a part of overall mental health. Young men viewed mental health as being stable enough to care for one's self. Older people were more likely to describe mental health as an illness, or, in the case of older women, as a set of health services that treat mental illness.

3.3.3.2 *Well-being*

Well-being was viewed by most groups as being able to keep a balance in life, reflecting people's day-to-day physical and emotional state and their ability to control their emotions. The female groups described well-being as an interaction of both physical and mental health. The general men's group also noted a distinction between mental and physical well-being although recognised that one could affect the other:

"Yeah, and you can have, you can have, eh, emotional symptoms of a physical condition or the opposite way. You can have physical symptoms relating to a mental health problem."

(Women socio-economic groups C2, D, E)

"I think that one comes with the, with the other things. They're interlinked but some symptoms are stronger than others depending on the individual and different times."

(Women socio-economic groups C2, D, E)

"Some of it... it would be part of emotional mental health, but it doesn't talk about any of the physical aspects of stress, or things like outside factors that... You know, like that can be connected with it."

(Women 25-64)

3.3.3.3 *Mental illness and mental well-being*

Participants were asked to describe whether they thought that mental illness and mental well-being could co-exist. There appeared to be a consensus (with the exception of the older women's groups) that if the individual's difficulties were managed properly then they could experience well-being.

The young men's group also identified that states of mental well-being could be transient and that events could trigger a shift into emotional distress:

"You can't... you can just go out and have fun but you're still going to be thinking about the problems you've got and then you're going to stop and think and then you're going to bring yourself down and while you're having fun you're still going to bring yourself down by thinking about it. Something is going to bring up that problem, something mentioned or something that's been said is going to bring that problem up and you'll find it's just going to go bang!"

(Men 18-24)

NOT FOR WIDER CIRCULATION OR REPRODUCTION

3.3.4 *Extra items*

After a general discussion and clarification about the purpose of the scale, participants were asked whether there were areas that affected their mental health that were not covered by Affectometer 2.

The general women's group noted that Affectometer 2 does not include items describing physical manifestations of mental state such as being able to weep. They suggested adding items relating to specific mood-states: 'feeling irritable' and 'feeling angry' and more items addressing an individual's relationships with others.

The group of older women suggested that the concept of feeling 'trapped', and a statement relating to relationships should be included.

3.3.5 *Sentences versus words*

Participants were asked whether they found the statement or word items easier to complete. Although there was no clear consensus, several participants noted that the statements section made more immediate sense, whereas the words section required the individual to personalise each item before they could complete it:

"Instead of just saying 'lonely' and 'good-natured', [say] 'do you feel good-natured', just tae (to) bring it close tae (to) yourself a wee (little) bit."
(Women 18-24)

"Because the words don't...the word probably don't make you feel there's emotion attached to them, because the statements say 'I', so they make it very...the statements personalise it where the words don't."
(Women 25-64)

"Yes and it's like 'I'm a failure.' You need to put 'I feel like a failure', or 'I feel unlike myself' or something like that, rather than just saying 'like a failure.'
(Men 25-64)

3.3.6 *Differences between focus groups results*

The most obvious difference between results of the focus groups was the difficulty older people had in completing Affectometer 2. The older men and older women's groups both had difficulties following each item along the line to record their response in the correct place. Two older participants also had reading problems and had to be helped to complete the scale by either the focus group moderator or the researcher. Some older people also had difficulty responding to how they felt over a two week period, because of failing memory.

There did not appear to be identifiable differences in how the groups responded to the scale, or how they conceptualised mental health. There appeared to be a slight tendency for older people, particularly older women, to equate mental health with mental illness and to not accept that mental health could also refer to a more positive state. However, this theme emerged in most of the focus groups, suggesting that the concept of a positive dimension to mental health is not widely recognised.

NOT FOR WIDER CIRCULATION OR REPRODUCTION

Women's groups were more likely than the men's groups to identify items or moods that they thought affect their mental health that were not covered by the scale. Both the older women's group and the general women's group felt that the scale did not cover relationships with other people adequately.

3.3.7 Focus group participants

The focus group participants were mostly recruited in areas of high deprivation. Our analysis may therefore not capture the views of higher socio-economic groups. Previous research has shown that low socio-economic groups may be less likely to complete questionnaires like Affectometer 2. It may therefore be more important to understand the views of these groups than of higher socio-economic groups.

Several of the focus group participants (excluding those in the mental health service user group) had experienced mental illness: the lower socio-economic status men's group were recruited from a project set up to help men to cope with mental health problems, and several women in the general women's groups had experienced episodes of clinical depression. Our analysis suggests that participants who had overcome, or were living with mental health problems were more likely to be familiar with concepts of positive mental health or well-being than those participants who did not have direct experience of mental illness.

3.4 Key conclusions

The focus groups showed that Affectometer 2 appears to have good face validity across a range of socio-economic groups. With the exception of older people, participants generally found the scale easy to complete. The problems older people reported would be common to any scale and suggests that extra care needs to be taken when using self report scales with this age group.

Participants struggled to come up with a clear definition of what the scale was measuring but their descriptions were broadly consistent with current concepts of positive mental health. However, although Affectometer 2 attempts to measure positive and negative dimensions of affect in equal proportions, this was not recognised by any of the participants. Several participants felt that the scale was overwhelmingly negative.

The sentence component of the scale appeared to be easier to understand than the word component. Individuals identified specific words or concepts that they thought the scale could incorporate. All the new items identified referred to negative aspects of mental health and there was no clear consensus about which items should be added.

The analysis of this phase of the project suggested a number of possible amendments that could be made to Affectometer 2 as part of the process of developing a shorter scale (WEMWBS).

- amend the Likert scale to make distinction between response categories clearer
- develop a revised version that uses sentences in place of a combination of words and sentences

NOT FOR WIDER CIRCULATION OR REPRODUCTION

- consider focusing on positive aspects of mental health.

These amendments are discussed in more detail in section 5: Phase 4: Development of the Warwick-Edinburgh Mental Well-being Scale.

4 Phase 3: Psychometric testing of Affectometer 2 with student samples

4.1 Aims

The purpose of this phase of the project was to test Affectometer 2 with respect to:

- construct validity (whether the scale shows anticipated relationship to other scales measuring aspects of mental health and well-being)
- response bias (the extent to which people respond in a way that presents them in a positive, or negative light)
- internal consistency (whether the items in the scale describe a consistent underlying theme)
- test-retest reliability (stability over time)
- performance of the four sub-scales (positive items, negative items, statements, words) compared to the performance of the full scale;
- comparison with data collected in the Scottish Health Education Population Survey (HEPS), 2002.

This phase of research was carried out from November 2005 to January 2006.

4.2 Method

4.2.1 Ethics approval and consent

Approval for the study was given by Warwick Medical School Ethics Committee. Students from Warwick and Edinburgh Universities were given an information sheet about the study (Appendix 6) one week before the questionnaires were distributed to decide whether or not they wanted to take part in the study. Those who did were asked to sign two consent forms (Appendix 7) (one for themselves and one for the project researcher).

4.2.2 Sample size & questionnaire pack administration

A sample size of 800 split equally between Warwick and Edinburgh was calculated to allow for analysis of data by study-site, assuming a 50% response rate. For assessing test-retest reliability, a sample size of 300 students for the second week of testing was calculated, assuming a 50% response rate. The sample size calculation for both weeks of testing was based on the number of responses required to carry out factor analysis on the dataset. Further details are available from the report authors.

Student groups were selected to maximise the variability between the samples and included post-graduate and undergraduate students from different disciplines. Details of these groups are given in Table 11 (week 1) (section 4.3.2). Students were recruited by approaching tutors to ask for permission for the researchers to attend three consecutive lectures or seminars over a three-week period. Information sheets were distributed to all students in the first week and questionnaire packs were administered at the end of formal teaching time in the second and third weeks.

NOT FOR WIDER CIRCULATION OR REPRODUCTION

Participating students were each given a pack containing two consent forms, a questionnaire pack containing Affectometer 2 (with either the statements or words first) and other questionnaires (Table 7), and an envelope addressed to the researcher administering the process at each site, to be returned using the University’s internal mail system. The questionnaire pack included a demographic data sheet asking students for their age, sex, ethnicity and a unique identifier. The unique identifier was used to match responses from students in week 1 with responses from those students who were asked to complete a second pack one week later (see below). Depending on whether there was time available when the packs were handed out, students either completed the questionnaires on the spot or took them away, completed them in their own time and returned them to the site researcher using the envelope provided. Recruitment of students and questionnaire pack administration was carried out by one researcher at each site.

Table 7: Questionnaire pack contents

Pack	Content
Version 1a	Affectometer 2 (statements first), Balanced Inventory of Desirable Response, Satisfaction with Life Scale, Positive and Negative Affect Scale, Global Life Satisfaction
Version 1b	Affectometer 2 (words first), Balanced Inventory of Desirable Response, Satisfaction with Life Scale, Positive and Negative Affect Scale, Global Life Satisfaction
Version 2a	Affectometer 2 (statements first), Global Life Satisfaction, Scale of Psychological Well-being (Ryff), WHO-5, Short Depression Happiness Scale
Version 2b	Affectometer 2 (words first), Global Life Satisfaction, Scale of Psychological Well-being (Ryff), WHO-5, Short Depression Happiness Scale
Version 3a	Global Life Satisfaction, Affectometer 2 (statements first), Emotional Intelligence Scale, EQ-5D
Version 3b	Global Life Satisfaction, Affectometer (words first), Emotional Intelligence Scale, EQ-5D

One week later, the researchers attended the same student group teaching sessions and randomly handed out the demographic data sheet and Affectometer 2 to one third of the students who had attended the previous week’s teaching sessions. Responses from this sub-sample were matched to the student’s responses from the previous week using the unique identifier. The number of students and composition of the student groups is set out in Table 12 (section 4.3.2).

4.2.3 Assessing construct validity

Construct validity was assessed by testing correlations between Affectometer 2 and eight other scales (Table 8).

NOT FOR WIDER CIRCULATION OR REPRODUCTION

Scales were selected to cover the key aspects of positive mental health described in the literature: positive affect/feelings, life satisfaction and positive psychological functioning. A scale of emotional intelligence was also included. Emotional intelligence (sometimes referred to as emotional literacy or emotional competence) is represented in the academic literature, largely from the USA, as a distinct concept unrelated to positive psychology or subjective well-being (Salovey and Mayer, 1990). However, it is widely regarded as a key skill for interpersonal relationships which are part of the positive psychology perspective. Poor interpersonal relationships are also recognised as part of the picture of poor mental health. Emotional intelligence is represented in UK policy relating to children and young people as a key skill for mental health (Department for Education and Skills, 2004, Department of Health 2004). The EQ-5D Thermometer was also included as a scale of general health.

Table 8: Scales used to assess Affectometer 2's construct validity

Scales of affect/feelings

WHO Well-being Index (WHO-5) - Five item scale of statements covering key mental affect states (e.g. I felt cheerful, calm, vigorous and interested) with 5 response categories. All items positively worded.

Short Depression Happiness Scale (SHDS) - Six item scale with 4 response categories focused on affect with balanced positive and negative items (e.g. I felt happy, I felt cheerless).

Positive And Negative Affect Scale (PANAS) - Twenty item scale with 5 response categories comprising a list of positive and negative adjectives covering a wider variety of feelings than is usual in mental health scales (e.g. ashamed, attentive, proud, guilty, and excited).

Scales of subjective well-being

Global Life Satisfaction Scale - Single item scale with 4 point response category. 'On the whole are you satisfied with your life?' Most commonly used measure of subjective well-being.

Satisfaction with Life Scale (SWLS) - Five item scale with 7 response categories. Items cover positive statements e.g. 'in most ways my life is close to ideal'. The prototype measure of well-being.

Psychological functioning

Scale of Psychological Well-Being - Fifty four item scale with six response categories assessing psychological functioning with subscales measuring autonomy, self-acceptance, environmental mastery, purpose in life, personal growth and positive relations with others.

Emotional Intelligence

Emotional Intelligence Scale - Thirty three item scale with five response categories. Consists of statements covering appraisal, expression, and regulation of emotion in self and others, and the utilisation of emotions in problem solving.

General health

EQ-5D thermometer - asks respondents to rate their overall health (i.e. physical as well as mental) on a 0-100 scale. Responses to this scale tend to reflect physical more than mental health.

Prior hypotheses were developed, providing expected correlations between Affectometer 2 and each of the scales, against which its performance was tested. These hypotheses were developed by the project Advisory Group using a consensus process (Table 9). Given the absence of clear differences in definition between alternative concepts of positive mental health, well-being and happiness, it was not possible to generate precise expected correlations between scales. Equally, there are no widely-accepted empirical guidelines for interpreting the strength of association between correlation coefficients (Hemphill, 2003). Cohen (Cohen 1988) suggests that a correlation of 0.5 should be regarded as large, but as all the scales selected aim to measure concepts related to those measured by Affectometer 2, we anticipated all scales to be correlated >0.5. We therefore adopted more stringent criteria as set out in Table 10.

Table 9: Predicted correlations between Affectometer 2 and other scales

Scale	Hypothesised relationship
Satisfaction with Life Scale	Low positive correlation. Scale focuses on cognitive aspects of satisfaction rather than affect or functioning.
Global Life Satisfaction	Low positive correlation. Scale focuses on cognitive aspects of satisfaction rather than affect or functioning.
Emotional Intelligence Scale	Low positive correlation. Scale focuses on understanding of emotions in self and others and impact of emotions on psychological functioning.
PANAS- NA (negative sub-scale)*	Moderate negative correlation. Scale covers a wide range of negative feeling adjectives.
PANAS- PA (positive sub-scale)*	Moderate positive correlation. Scale covers a wide range of positive feeling adjectives.
EQ-5D	Moderate positive correlation reflecting global health content of measure and interplay between mental and physical health.

Scale	Hypothesised relationship
WHO-5	Moderate to high positive correlation. Statements reflect positive affect.
Scales of Psychological Well-being	Moderate to high positive correlation. Statements reflect a wide range of psychological functioning
Short Depression Happiness Scale	Moderate to high positive correlation. Scale reflects positive and negative affect.

* Scores for the two components of the PANAS scale are not aggregated into a single scale score.

Table 10: Interpretation of correlation coefficients

Correlation coefficient range	Interpretation
0.80 to 1.00	High correlation
0.60 to 0.79	Moderate correlation
<0.59	Low correlation

4.2.4 Relationship between positive and negative items in Affectometer 2

The relationship between positive and negative items in the scale was assessed by correlating the sum of all positive item scores with the sum of all negative item scores, using Pearson's correlation coefficient.

4.2.5 Assessing test-retest reliability

Test-retest reliability was assessed by administering Affectometer 2 to the same students one week apart. Test-retest reliability aims to measure how stable a scale is over time. It was hypothesised that the scale would not show a very high or perfect correlation (>0.90) as responses would be partly affected by current state, which could be different at the two time points when the scale was administered. It was expected that the test-retest correlation would show a moderate to high correlation (between 0.7 and 0.9), reflecting an individual's longer-term trait.

4.2.6 Assessing response bias

Response bias was tested by collecting data from a sub-sample of participants using the Balanced Inventory of Desired Responding (BIDR). The BIDR is a 40-item scale, split into two sub-scales. The first sub-scale measures self-deception (the tendency to exaggerate certain responses or behaviours) and the second sub-scale measures impression management (the tendency to over-report desirable behaviours and under-report undesirable behaviours) (Robinson et al, 1991). It was hypothesised that, for Affectometer 2 to be an effective tool, it should show similar correlations with the BIDR to those shown by other scales tested in this research.

4.2.7 Data Entry and Quality

All data were entered by a clerical assistant at Warwick University. Random checks were carried out to assess the quality of data entry and showed an accuracy of 99.4%. Participants with incomplete scales were removed from the analysis,

4.2.8 Statistical tests

Mean scores for each scale were calculated for the whole sample for each site and for each subject group. The significance of differences in scores were calculated using t-tests (2 groups) or one-way ANOVA (>2 groups). Correlations between Affectometer 2 scores and each of the other scales were calculated using Pearson's correlation coefficient. Correlations between Affectometer 2 scores in week 1 and week 2 were calculated using intra-class correlation coefficients. Affectometer 2's internal consistency was assessed using Cronbach's alpha. Response bias was assessed by calculating Pearson's correlation coefficients between Affectometer 2 scores and the two scales of the Balanced Inventory of Desired Responding. All statistical tests were carried out using SPSS version 12.0.1.

4.3 Results

4.3.1 Data analysis

Eleven percent (n=56) of responses were incomplete and were omitted from the analysis phase.

4.3.2 Response rates

Response rates in week 1 were 62% across the whole sample, giving a total number of responses of 437, out of 710 and exceeding the number (400) estimated as desirable in the sample size calculation (Table 11).

Table 11: Week 1 response rates %

University	Subject	Number given packs	Number who completed packs	Response rate
Warwick	Philosophy (UG)	127	70	55%
Warwick	Medicine (PG)	173	146	96%
Warwick	Education (PG)	65	43	66%
Edinburgh	Medicine (UG)	89	85	96%
Edinburgh	Social Policy (UG)	27	13	48%
Edinburgh	Social Science (PG)	76	36	47%
Edinburgh	Psychology (UG)	153	44	29%
Total		710	437	62%

UG= undergraduate, PG=postgraduate

The week 2 response rate was 74%, giving a total number of responses of 203 out of a total of 274, exceeding the estimated necessary sample size of 150 (Table 12).

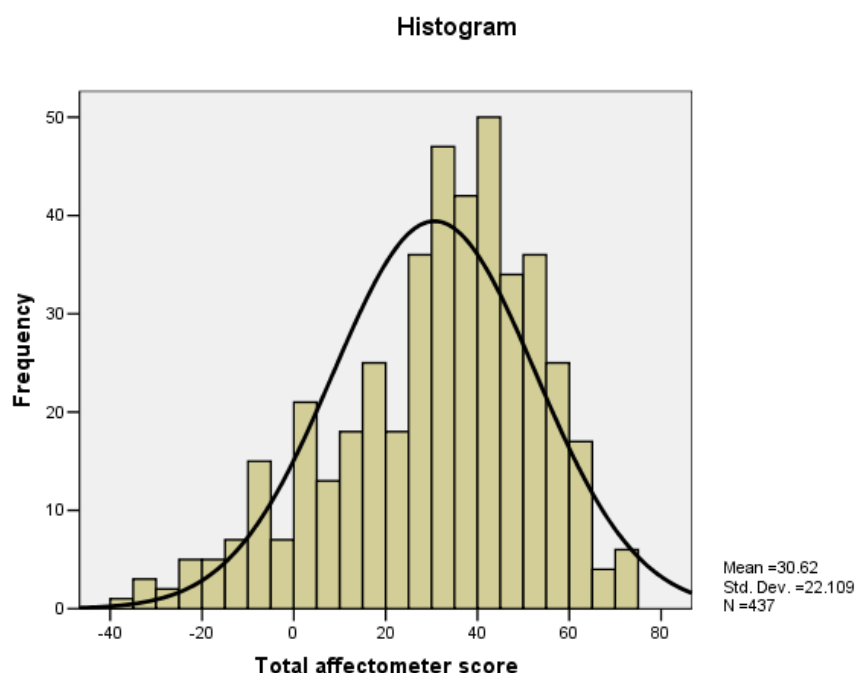
Table 12: Week 2 response rates %

Site	Subject	Number given packs	Number who completed packs	% of respondents who completed both packs
Warwick	Philosophy	53	35	66%
Warwick	Medicine	56	55	98%
Warwick	Education	24	17	71%
Edinburgh	Medicine	50	43	86%
Edinburgh	Social Policy	8	6	75%
Edinburgh	Social Science	35	24	69%
Edinburgh	Psychology	48	23	48%
Total		274	203	74%

4.3.3 Affectometer 2: descriptive data

Affectometer 2 scores followed a roughly Normal distribution, indicating that the scale does not show ceiling effects. Scores also used most of the available score range (Figure 1).

Figure 1: Distribution of Affectometer 2 scores



NOT FOR WIDER CIRCULATION OR REPRODUCTION

Mean scores and score ranges for the full scale were similar to the scores observed in the analysis of population data collected in the 2002 Health Education and Population Survey (mean score, 29.67) although the range of scores was narrower (-38 to 74 compared with -59 to 77 in HEPS) (see Appendix 2: Analysis of HEPS dataset).

Table 13: Mean scores and sub-scale scores for Affectometer 2

	Mean score	SD	Range of scores	95% Confidence Intervals
Full scale	30.62	22.11	-38 to 74	28.55 to 32.70
Words	14.32	11.08	-22 to 38	35.84 to 36.76
Statements	16.50	11.60	-25 to 39	15.45 to 17.56
Positive items	51.15	11.74	20 to 78	50.07 to 52.23
Negative items	20.36	11.85	1 to 62	19.27 to 21.46

There were no significant differences in mean scores for the full scale by age-group (one-way ANOVA) ($p=0.294$) (Table 14). This result contrasts with the results of analysis of HEPS data which found significant differences between scores based on age and may reflect the narrow range of ages in this sample of students (see Appendix 2: Analysis of HEPS dataset).

Table 14: Mean Affectometer 2 scores by age (whole sample)

Age group	Number of respondents	Mean score	SD	Range of scores	95% Confidence intervals
18-22	267	30.09	22.35	-38 to 74	27.39 to 32.78
23-27	107	33.86	21.05	-21 to 71	29.82 to 37.90
28+	61	27.89	22.72	-38 to 71	22.07 to 33.70

The difference in mean scores between men and women was not statistically significant (independent samples, t-test) ($p=0.151$) (Table 15). This differs from the findings from the analysis of data collected in HEPS (see Appendix 2: Analysis of HEPS dataset) where the scores of men were higher than those of women.

Table 15: Mean Affectometer 2 score by sex (full scale)

Sex	Number of respondents	Mean score	SD	Range of scores	95% confidence intervals
Male	130	33.55	20.50	-32 to 70	29.99 to 37.10
Female	297	29.49	22.73	-38 to 74	26.89 to 32.08

Mean scores among Warwick and Edinburgh students were not significantly different (independent samples t-test) ($p=0.790$) (Table 16). There was however a significant difference in Affectometer 2 scores among the seven subject groups ($p<0.01$), (one-way ANOVA). The mean score for philosophy students (26.16) was significantly lower than that that for medical students (34.92) ($p<0.01$) and mean score for medical students was higher than that of both social policy students (13.31) and psychology students (22.98) ($p<0.01$).

Table 16: Affectometer 2 score by student sample

Sample	Number of respondents	Mean score	SD	Range of scores	95% Confidence intervals
Whole sample	437	30.62	22.11	-38 to 74	28.55 to 32.70
Warwick (all students)	259	31.34	21.86	-32 to 74	28.67 to 34.02
Edinburgh (all students)	178	29.58	22.49	-38 to 74	26.25 to 32.90
Warwick philosophy	70	26.16	21.18	-32 to 63	21.11 to 31.21
Warwick medical students	146	34.71	21.47	-29 to 74	31.20 to 38.22
Warwick education	43	28.35	22.65	-26 to 63	21.38 to 35.32
Medical students (overall)	131	34.92	20.64	-38 to 74	32.25 to 37.60
Edinburgh Medicine	85	35.28	19.23	-38 to 70	31.13 to 39.43
Edinburgh Social Policy	13	13.31	29.58	-35 to 71	-4.57 to 31.18
Edinburgh Social Science	36	30.06	19.42	-18 to 64	23.49 to 36.63
Edinburgh Psychology	44	22.98	24.88	-35 to 63	15.41 to 30.54

NOT FOR WIDER CIRCULATION OR REPRODUCTION

Scores on the version of Affectometer 2 where the word items came before the statements were significantly higher than those on the version that started with sentence items (Table 17) (independent samples t-tests).

Table 17: Score differences by order of statements and words

Order	Mean score	p-value
Words first	34.08 (SD=20.51)	0.02
Statements first	27.42 (SD=23.08)	

4.3.4 Internal consistency

Internal consistency, as measured by Cronbach’s alpha was high, both for the full scale and each of the four potential sub-scales (Table 18). This is likely to reflect the large number of items in the scales and sub-scales and indicates item redundancy. These alpha coefficients are consistent with an analysis carried out on the HEPS dataset (see Appendix 2).

Table 18: Internal consistency of full scale and sub-scales

	Full scale	Statements	Words	Positive items	Negative items
Alpha Coefficient	0.95	0.91	0.92	0.93	0.92

4.3.5 Construct validity

For the full scale, the lowest correlations were observed with the EQ-5D and the Emotional Intelligence Scale (Pearson’s correlation coefficient) (Table 19). Moderate correlations were observed with the two life satisfaction scales and with both the positive and negative PANAS sub-scales and the Short Depression Happiness Scale. WHO-5 and the Scales of Psychological Well-Being showed high correlations. Although the negative items sub-scale showed lower correlations than those for the full scale score, similar patterns were observed for the four sub-scales. These correlations fulfilled some of the predictions relating to construct validity (Table 9). The exceptions were the life satisfaction scales which correlated somewhat better than expected, and the EQ-5D and the Short Depression Happiness Scale which correlated somewhat worse than expected. Of these, the Short Depression Happiness scale correlated at the same level as other scales of affect/feelings (PANAS positive and negative sub-scales) and of life satisfaction. As a measure that contains a mixture of life satisfaction and affect items, in retrospect, it seems reasonable that we observed this level of correlation between the Short Depression Happiness Scale and Affectometer 2 and, in retrospect some predictions were probably inappropriate. The lower correlation with the EQ-5D, primarily a scale of physical health, is also acceptable. The high correlation with scales of both Psychological Well-being (covering functioning) and the WHO-5 (covering primarily positive affect/feelings) suggests that Affectometer 2 is covering both functioning (eudaimonic perspective) and affect (hedonic perspective).

Table 19: Correlations between Affectometer 2 (full and sub-scale) score and other scales

Scale	Expected correlation	Full scale	Positive items	Negative items	Statements	Words
Satisfaction with Life Scale	0.40 to 0.59	0.64**	0.65**	-0.51**	0.66**	0.58**
Global life satisfaction	0.40 to 0.59	0.63**	0.62**	-0.52**	0.61**	0.59**
Emotional Intelligence Scale	0.40 to 0.59	0.50**	0.54**	-0.38**	0.48**	0.47**
EQ-5D	0.60 to 0.79	0.44**	0.40**	-0.43**	0.42**	0.45**
PANAS- NA	0.60 to 0.79	-0.62**	-0.48**	0.69**	-0.61**	-0.59**
PANAS- PA	0.60 to 0.79	0.60**	0.70**	-0.37**	0.55**	0.59**
WHO-5	0.75 to 0.85	0.81**	0.79**	-0.75**	0.80**	0.78**
Scales of Psychological Well-being	0.75 to 0.85	0.80**	0.78**	-0.73**	0.81**	0.77**
Short Depression Happiness scale	0.75 to 0.85	0.59**	0.57**	-0.53**	0.54**	0.59**

**significant at 0.05 level*

*** significant at 0.01 level*

4.3.6 Relationship between positive and negative items of Affectometer 2

The correlation between positive and negative item scores was -0.748 ($p < 0.01$) (Pearson's Correlation Coefficient) suggesting a strong but not perfect link. So while people with high positive scores were likely to also have low negative scores, the level of correlation suggests that the two are not a mirror image of each other.

4.3.7 Test-retest reliability

Test-retest reliabilities were assessed using intra-class correlation coefficients, for the 203 students who completed Affectometer 2 one week later. The correlations were high for the full scale and four sub-scales (Table 20), falling within the hypothesised range of 0.7 to 0.9. Minor differences were observed between the Warwick and Edinburgh samples. The statement sub-scale showed the lowest level of test-retest reliability of the four sub-scales for the whole sample and for the Edinburgh sample, but the level was high even for this scale. These results indicate that Affectometer 2 is tapping into a longer-term trait and does not appear to be susceptible to more transient changes in mental state.

Table 20: One-week test-retest reliability of Affectometer 2

	Full scale	Positive items	Negative items	Statements	Words
Whole sample	0.84**	0.84**	0.82**	0.81**	0.84**
Edinburgh	0.77**	0.76**	0.76**	0.72**	0.77**
Warwick	0.90**	0.89**	0.87**	0.88**	0.89**

**significant at 0.05 level*

*** significant at 0.01 level*

4.3.8 Response bias

Mean scores for the two sub-scales of the Balanced Inventory of Desirable Responding (BIDR) (impression management and self-deception) were 6.6 (SD=2.9) for men and 5.3 (SD=2.7) for women for impression management and 5.7 (SD=3.0) for men and 4.7 (SD=2.7) for women for self-deception. These scores were lower than the range reported in U.S. samples for self-deception (7.6 to 7.6 for men, 6.8 to 7.3 for women) but comparable for impression management (4.3 to 7.3 for men, 4.9 to 8.9 for women) (Robinson et al, 1991). This suggests that our student sample had a lower tendency to self-deception than the general population (at least in the US).

The full scale and four subscales showed a very similar pattern of correlation with the BIDR scales (Pearson's correlation coefficient) (Table 21).

Correlations with the impression management sub-scale of the BIDR were low and in all cases (except the negative item sub-scale) negative, indicating that people with high Affectometer 2 scores were likely to have low levels of impression management i.e. were likely to be being more honest.

Table 21: Correlations between Affectometer 2 scores and BIDR sub-scale scores

Affectometer	Impression management	Self-deception
Full scale	-0.251**	0.546**
Positive items	-0.211*	0.484**
Negative items	0.251**	-0.515**
Sentences	-0.281**	0.509**
Words	-0.212*	0.537**

**significant at 0.05 level*

*** significant at 0.01 level*

As Table 22 shows, Pearson's correlation coefficients between Affectometer 2 and the BIDR self-deception sub-scale were higher than for any other scale, suggesting

that Affectometer 2 is more prone to self-deception bias. This is a significant weakness. Correlations between the BIDR impression management sub-scale for the Affectometer 2 were similar to those for other measures.

Table 22: Correlation between BIDR sub-scale scores and some of the other scales used to test Affectometer 2

Scale	Impression management	Self-deception
Affectometer (full scale)	-0.251**	0.546**
Satisfaction with life scale	-0.270**	0.328**
Global life satisfaction	-0.218*	0.267**
PANAS –PA	-0.006	0.301**
PANAS – NA	0.293**	-0.394**

**significant at 0.05 level*

*** significant at 0.01 level*

4.4 Discussion

This round of testing showed a similar pattern and distribution of scores to those observed in a previous analysis of HEPS data (Appendix 2). Scores followed an approximately Normal distribution, indicating that Affectometer 2 is capable of capturing a full range of states of positive mental health. In common with the HEPS analysis, the scale was able to discriminate between different population groups, with differences emerging between students taking different subjects. Unlike analysis of the HEPS data, the scale did not discriminate between groups on the basis of age or gender; the first is likely to be due to the smaller age-range in this sample, and the second may be due to the fact that all were Higher Education students and the smaller number of men in the sample.

4.4.1 Internal consistency

As with the HEPS data analysis, the scale and the four sub-scales showed very high levels of internal consistency. There are no clear standards for an appropriate level of internal consistency; this is in part because internal consistency scores are influenced by the number of items in a scale, with longer scales showing higher scores (Schmitt, 1996). Cut-off values of 0.7 have been suggested (Nunnally, 1978) as indicating a suitable level of internal consistency. However, very high values may indicate redundancy within a scale (Boyle, 1991). An effective scale therefore needs to lie within an optimal range that is neither too low (indicating poor internal consistency) nor too high (indicating item redundancy). In the case of Affectometer 2, the internal consistency score was similarly high (0.91 to 0.95) for both the 40 and 20-item versions of the scale. This indicates that the number of items in the scale could be reduced to below twenty, which would also make the scale easier and less time-consuming to complete.

4.4.2 Construct validity

The scale appears to have good construct validity, showing high correlations with other scales that measure aspects of positive mental health, such as WHO-5 and the Scales of Psychological Well-being. It also showed lower correlations with scales that measure related but different concepts such as general life-satisfaction, overall health and emotional intelligence. This suggests that the scale is capturing a distinct construct of positive mental health which includes psychological functioning as well as affect/feelings.

4.4.3 Test-retest reliability

The scale shows good test-retest reliability with mean scores staying relatively stable over the one week period between tests. This indicates that the scale is capturing general positive mental health (psychological trait), rather than a transitory mood (psychological state).

4.4.4 Response bias

The correlations between the Affectometer 2 full scales and sub-scales showed a higher than expected correlation on the self-deception sub-scale and a low correlation with the impression management sub-scale. Ideally, Affectometer 2 should show no, or a low correlation with both BIDR sub-scales, indicating that people's responses to the scale are not conditioned by how they want to appear (impression management) or by unrealistic perceptions of their own behaviours or beliefs (self-deception). On the basis of BIDR scores, people with higher Affectometer 2 scores were less likely to try and present themselves in a more positive light than people with a low Affectometer 2 score. In contrast, people with a high Affectometer 2 score were more likely to have higher self-deception scores than people with a low Affectometer 2 score. Both of these phenomena were also observed for the other scales against which the BIDR was tested.

A key problem with the Affectometer 2 scale is that it appears to be more subject to certain forms of response bias than other comparable scales. While most of the scales tested (including Affectometer 2) showed relatively high susceptibility to impression management, Affectometer 2 showed the highest score of any scale tested for self-deception bias.

4.5 Key conclusions

Two main issues emerge from this phase of research: first, although the scale shows good construct validity and test-retest reliability, its high internal consistency points to the potential for shortening the scale. Second, the impact of self-deception bias would need to be closely examined in any revised version of Affectometer 2.

5 Phase 4: Development of the Warwick-Edinburgh Mental Well-being Scale

5.1 Aims

This phase of the research built on the previous phases with the aim of producing a shorter scale that was likely to be both clear and meaningful to adults from different social backgrounds in the UK.

5.2 Method

Several methods were used to make revisions:

- identification of items that could be removed from Affectometer 2 without affecting the scale's internal consistency
- identification and removal/modification of items with completion rates <99%
- identification of the number of items overall which could be removed without reducing the Cronbach's alpha below 0.8
- identification of items that focus group participants found confusing or difficult to respond to, for example the adjective scales were identified as more difficult to respond to than the statement scales. Cronbach's alpha scores for the scale with each of these items removed were also examined
- referring back to Affectometer 2 to ensure that key concepts were still covered
- referring back to the current positive mental health literature to check to that key concepts were still covered.

5.3 Results

5.3.1 Removing scale items

As Table 23 shows, around thirty items could be eliminated from the scale before the Cronbach's alpha dropped below an acceptable limit of 0.80.

Table 23: Cronbach's alpha for different number of items remaining in Affectometer 2

Number of items	Minimum Alpha Coefficient	Maximum Alpha Coefficient
30	0.937	0.942
20	0.907	0.915
15	0.881	0.889
10	0.857	0.863
5	0.774	0.795

NOT FOR WIDER CIRCULATION OR REPRODUCTION

5.3.2 New scale development

A draft fourteen item scale comprising both positive and negative worded items and a revised Likert response scale was presented to the project Advisory Group for comment following the six steps method above. There was agreement that this version was concise and unambiguous, with good face validity and an adequate internal consistency based on data collected on Affectometer 2 in phase 3.

Key points discussed included the importance of including items relating to personal development and purpose in life/spirituality (represented in Affectometer 2 by items 'it seems as though the best years of my life are over' and 'my life seems stuck in a rut'), and autonomy (represented in Affectometer 2 by the item 'I can do whatever I want to do').

The balance of positive and negative worded items was also debated. There was clear agreement that the scale needed to reflect positive attributes of mental health, but also concern to avoid response bias induced by phrasing all items positively.

In the event, drafts of negatively worded statements resulted in double negatives, when combined with the Likert response scale, which were hard to make sense of, for example "My thoughts don't go round in useless circles – none of the time". For this reason, a decision was taken that the scale would comprise of only positively worded statements reflecting only positive aspects of mental health.

Further iterations of the scale were developed and commented on by the Advisory Group by email.

A new item reflecting interest in learning new things was drafted to represent personal development. Various statements were also drafted to represent autonomy as the original item in Affectometer 2 was deemed inappropriate. The chosen item ('I've been able to make up my own mind about things') represented only one component of autonomy. Various positively worded items relating to purpose in life or spirituality were considered, but none was thought suitable for general population surveys. Although the group were clear that this aspect of positive mental health was important, the decision was made to leave items related to spirituality/purpose in life out of the final scale on the grounds that this is not accepted as important for mental health in all population groups and might cause response problems.

The final scale retained a preponderance of items relating to interpersonal relationships. This was deemed appropriate by members of the Advisory Group on the grounds that academics writing from different perspectives on positive mental health and members of the public taking part in the focus groups all seem to agree on the importance of positive interpersonal relationships.

5.4 Conclusion - Warwick-Edinburgh Mental Well-Being Scale

The final scale, named the Warwick-Edinburgh Mental Well-being Scale (WEMWBS), consisted of 14 statements (Table 24) covering both hedonic and eudaimonic aspects of positive mental health, including positive affect (feelings of optimism, cheerfulness, relaxation), satisfying interpersonal relationships and positive functioning (energy, clear thinking, self acceptance, personal development, mastery

NOT FOR WIDER CIRCULATION OR REPRODUCTION

and autonomy). It does not include a specific item covering life satisfaction, but hedonic well-being is well represented. All items represent the positive end of the spectrum of mental health and are phrased positively. Individuals completing the scale are required to tick the box that best describes their experience of each statement over the past two weeks using a 5-point Likert scale (none of the time, rarely, some of the time, often, all of the time). The Likert scale represents a score for each item from 1 to 5 respectively (see Appendix 8). The overall score for the WEMWBS is calculated by totalling the scores for each item. This differs from the scoring of Affectometer 2 which had both positive and negative words and statements, requiring a more complicated calculation in comparison to the current scale.

Table 24: Warwick-Edinburgh Mental Well-Being Scale items

I've been feeling optimistic about the future
I've been feeling relaxed
I've had energy to spare
I've been thinking clearly
I've been feeling close to other people
I've been able to make up my own mind about things
I've been interested in new things
I've been feeling useful
I've been feeling interested in other people
I've been dealing with problems well
I've been feeling good about myself
I've been feeling confident
I've been feeling loved
I've been feeling cheerful

6 Phase 5: Psychometric testing of the Warwick-Edinburgh Mental Well-being Scale with student samples

6.1 Aims

The purpose of this phase of the research was to test the Warwick-Edinburgh Mental Well-being Scale with student samples with respect to:

- construct validity
- internal consistency
- test-retest reliability
- response bias

Testing took place between February and March 2006.

An additional phase of research was also carried out to test the correlation between Affectometer 2 and WEMWBS.

6.2 Method

The methodology (sample size calculation, sampling method, questionnaire pack administration & statistical methods) for this phase of testing with students was identical to that used during Phase 3:- Psychometric testing of Affectometer 2 with student samples (section 4.2).

For assessing the correlation between Affectometer 2 and WEMWBS a convenience sample of 72 adults from Scotland and England was used.

6.2.1 Questionnaire Packs

All participants were given a pack containing the WEMWBS (Appendix 8) and between one and four of the scales used in the testing of Affectometer 2 (Table 25).

Table 25: Questionnaire pack contents

Pack	Content
Version 1	WEMWBS, Balanced Inventory of Desirable Response, Global Life Satisfaction, Satisfaction with Life Scale, WHO-5.
Version 2	WEMWBS, Balanced Inventory of Desirable Response, Positive and Negative Affect Scale.
Version 3	WEMWBS, Scale of Psychological Well-being (Ryff).
Version 4	WEMWBS, Emotional Intelligence Scale.
Version 5	WEMWBS, Short Depression Happiness Scale, EQ-5D.

The convenience sample was asked to complete both Affectometer 2 and WEMWBS, the order of which was varied.

6.2.2 Construct validity

It was hypothesised that the correlations between WEMWBS and the other scales used to test construct validity (see Table 9) would be similar to those observed for Affectometer 2 (Table 19), as WEMWBS is a modification of Affectometer 2 and aims to capture the same underlying state (positive mental health).

6.2.3 Analyses

The same statistical methods were used in this phase as in phase 3 (section 4.2.8). In addition, an exploratory factor analysis (principal component analysis) was carried out using the Scree plot method to determine the number of underlying factors to the scale. Cronbach's alpha scores were also calculated for reduced item versions of the scale to identify at what point the Cronbach's alpha would drop below a value of 0.8 (the usual cut-off point for identifying a high level of internal consistency). Items were deleted from this scale at random, using three different options for each number of items. Pearson's correlation coefficient was used to test the correlation between Affectometer 2 and WEMWBS.

6.3 Results

6.3.1 Data analysis

Sixteen percent (n=58) of responses were incomplete and were omitted from the analysis phase.

6.3.2 Response rates

Response rates at week 1 were fifty three percent, giving a total of 354 responses, which did not quite achieve the recommended sample size (400) (However, in the event the estimated required sample size was an overestimate of that needed and the number of responses was sufficiently large enough, see Discussion) (Table 26).

Table 26: Week 1 response rates

University	Subject	Number given packs	Number who completed packs	Response rate
Warwick	Education (PG)	50	50	100%
Warwick	Statistics (UG)	51	23	45%
Warwick	Sociology (UG)	32	10	31%
Warwick	Econometrics (UG)	113	38	33%
Warwick	Medicine (PG)	96	20	21%
Edinburgh	Medicine (UG)	149	136	91%
Edinburgh	Social Policy (UG)	59	35	59%
Edinburgh	Politics (UG)	121	42	35%
Totals		671	354	53%

Response rates at week 2 were forty seven percent, giving a total of 124 responses, this was also below the recommended sample size (150) (as indicated above, the estimated required sample size was an overestimate of the sample size needed, and the number of responses was sufficiently large enough for statistical analyses, see Discussion) (Table 27).

Table 27: Week 2 response rates

Site	Subject	Number given packs	Number who completed packs	Response rate
Warwick	Education	20	15	75%
Warwick	Statistics	25	9	36%
Warwick	Sociology	20	6	30%
Warwick	Econometrics	35	5	14%
Warwick	Medicine	50	10	20%
Edinburgh	Medicine	50	44	88%
Edinburgh	Social Policy	19	17	89%
Edinburgh	Politics	47	18	38%
Totals		266	124	47%

6.3.3 Descriptive data

The WEMWBS shows a similar score distribution to Affectometer 2, with nearly the full range of scores being used (Figure 2). However, because the scoring systems used by the two scales are different, the mean from the WEMWBS is much higher and the standard deviation is smaller (Table 28).

Figure 2: WEMWBS score distribution

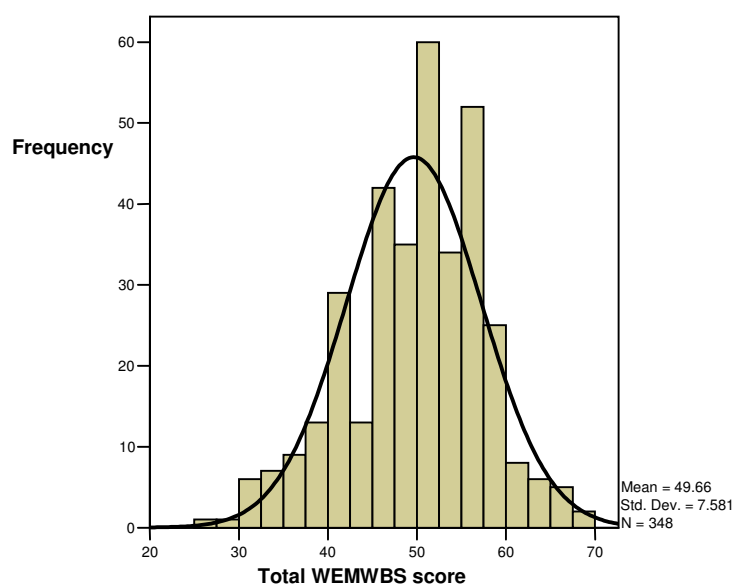


Table 28: WEMWBS scores (whole sample)

	Mean Score	SD	Range of scores	95% Confidence Intervals
WEMWBS	49.66	7.58	25 to 70	48.86 to 50.46

There were no significant differences in mean score by age group ($p = 0.105$) (one-way ANOVA) (Table 29). This finding is consistent with the results of the previous round of testing on Affectometer 2.

Table 29: Mean WEMWBS scores by age

Age group	No. of respondents	Mean score	SD	Range of scores	95% Confidence Intervals
18-22	275	49.78	7.51	25 to 70	48.89 to 50.67
23-27	36	50.75	8.20	33 to 67	47.98 to 53.52
28+	37	47.70	7.35	32 to 61	45.25 to 50.15

The mean scores for men and women did not differ (Table 30). This result contrasts with findings relating to Affectometer 2 in the HEPS data (Appendix 2: *Analysis of HEPS dataset*) but is similar to the findings based on the student sample (section 4.3.3). It also contrasts with findings on other measures of mental health, where in general women report worse mental health than men (Singleton, 2003).

Table 30: Mean WEMWBS scores by sex

Sex	No. of respondents	Mean score	SD	Range of scores	95% Confidence Intervals
Male	99	49.14	7.87	29 to 68	47.57 to 50.71
Female	244	49.87	7.43	25 to 70	48.94 to 50.81

Mean scores were significantly higher for Edinburgh students than Warwick students ($p < 0.01$) (independent samples, t-test) (Table 31). This is different to mean scores in the Affectometer 2 phase of testing (section 4.3.3).

Table 31: Mean WEMWBS score (whole sample and by university)

Sample	No. of respondents	Mean score	SD	Range of scores	95% Confidence Intervals
Whole sample	348	49.66	7.58	25 to 70	48.86 to 50.46
Warwick (all students)	139	47.73	7.33	29 to 64	46.50 to 48.96
Edinburgh (all students)	209	50.94	7.49	25 to 70	49.92 to 51.96

There was a significant difference in WEMWBS scores for the 8 teaching groups ($p < 0.01$) one to way ANOVA (Table 32). The mean score for education students (47.67) was significantly different from that for medical students (51.37) and the mean score for medical students (51.37) differed significantly from statistics students (46.05). A similar finding was observed in the first student sample, using Affectometer 2 (section 4.3.3).

Table 32: WEMWBS scores by subject

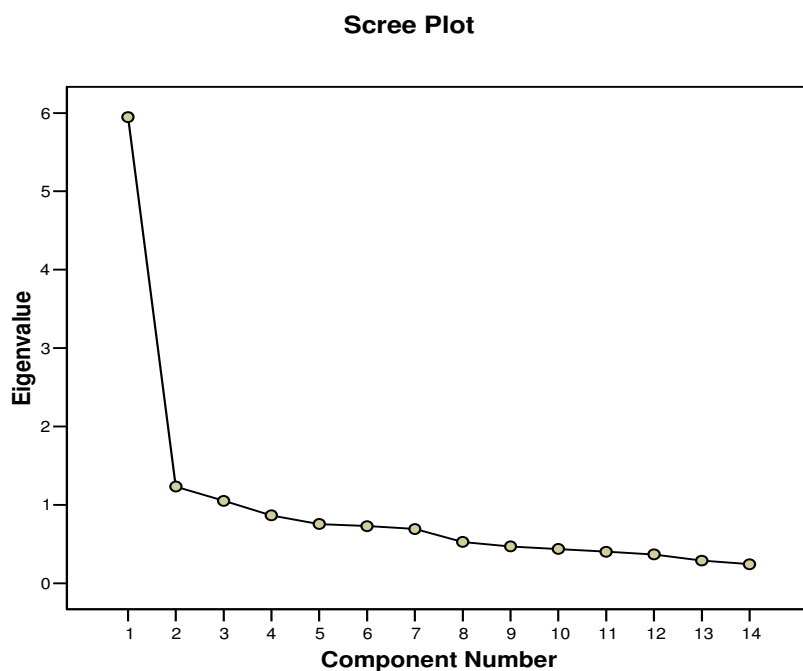
Sample	No	Mean	SD	Range of scores	95% Confidence intervals
Warwick Education	50	47.67	6.93	30 to 59	45.67 to 49.61
Warwick Statistics	21	46.05	8.26	29 to 62	42.29 to 49.81
Warwick Sociology	10	46.00	10.55	34 to 64	38.45 to 53.55
Warwick Econometrics	38	48.95	6.23	32 to 64	46.90 to 50.99
Medicine (whole sample)	153	51.37	7.49	25 to 70	50.17 to 52.56
Warwick Medicine	20	48.30	7.68	32 to 60	44.71 to 51.89
Edinburgh Politics	42	49.12	8.47	32 to 66	46.48 to 51.76

Sample	No	Mean	SD	Range of scores	95% Confidence intervals
Edinburgh Medicine	133	51.83	7.39	25 to 70	50.56 to 53.09
Edinburgh Social Policy	34	49.71	6.11	38 to 63	47.57 to 51.84

6.3.4 Factor analysis

Exploratory factor analysis (principal components analysis) was carried out to explore the number of dimensions in the revised scale. The Scree plot method (Figure 3) pointed to a single underlying factor which accounted for 42.5% of the total variance in the scale.

Figure 3: Scree plot for the WEMWBS



This analysis suggests that the revised scale represents a single underlying concept, making it appropriate to summarise items in a single score.

6.3.5 Internal consistency

The scale's internal consistency was tested using Cronbach's alpha (Table 33). Although the alpha coefficient was lower than for Affectometer 2, WEMWBS still demonstrates a high internal consistency. There are fewer items in WEMWBS, in comparison with the Affectometer 2 but the high internal consistency suggests that the number of items in WEMWBS could be reduced further.

Table 33: Internal consistency of WEMWBS

	WEMWBS
Alpha Coefficient	0.89

The results presented in Table 34 suggest that, the scale could be further reduced to around 9 items before the internal consistency falls below an acceptable limit of 0.8. It is not possible from this analysis to identify which items should be dropped.

Table 34: Internal consistency of WEMWBS versions with reduced numbers of items

Number of items	Minimum Alpha Coefficient	Maximum Alpha Coefficient
13	0.876	0.889
12	0.871	0.878
11	0.865	0.87
10	0.848	0.862
9	0.825	0.86
8	0.795	0.844
7	0.784	0.832

6.3.6 Construct validity

The construct validity of WEMWBS was tested by correlating it (Pearson's correlation coefficient) with the same scales (see Table 9) that were used in the validation of Affectometer 2 (Table 35).

Table 35: Correlations between Affectometer 2 and WEMWBS with other scales

Scale	Affectometer Correlation	WEMWBS Correlation
Satisfaction with Life Scale	0.64**	0.72**
Global life satisfaction	0.63**	0.55**
Emotional Intelligence Scale	0.50**	0.51**
PANAS- NA	-0.62**	-0.55**
PANAS- PA	0.60**	0.73*
WHO-5	0.81**	0.77**
EQ-5D thermometer	0.44**	0.42**

Scale	Affectometer Correlation	WEMWBS Correlation
Scales of Psychological Well-being	0.80**	0.73**
Short Depression Happiness scale	0.59**	0.76**

* significant at 0.05 level

** significant at 0.01 level

Moderate to low correlations were observed with the Global Life Satisfaction Scale, the Emotional Intelligence Scale, PANAS negative affect sub-scale and the EQ-5D. The PANAS positive affect sub-scale, Satisfaction With Life Scale, Scales of Psychological Well-being, WHO-5 and Short Depression Happiness Scale all showed high correlations.

The majority of correlations were similar to those observed with Affectometer 2. Correlations with the Satisfaction with Life Scale, and Short Depression Happiness Scale were increased over those observed with Affectometer 2.

6.3.7 Test-retest reliability

Test-retest reliabilities for WEMWBS for 1 week, assessed using the intra-class correlation coefficient, revealed a high reliability for WEMWBS which compared well with Affectometer 2's high reliability (Table 36).

Table 36: Test-retest reliability for WEMWBS

	Correlation for Affectometer 2	Correlation for WEMWBS
Whole sample	0.84**	0.83**
Warwick	0.90**	0.80**
Edinburgh	0.77**	0.83**

*significant at 0.05 level

** significant at 0.01 level

6.3.8 Response bias

Mean scores for the two subscales of the Balanced Inventory of Desirable Response (impression management and self-deception) were 6.7 (SD = 3.6) and 4.6 (SD = 3.2) respectively. There were no significant differences in mean impression management or self-deception scores between the different groups (one-way between subjects ANOVA). In comparison with Affectometer 2, WEMWBS had a lower correlation (Pearson's correlation coefficient) with both subscales, indicating that the new scale is less susceptible to biased response (Table 37). It also performed better than three comparison scales (except the PANAS positive and negative scales) on impression management and better than two the scales (excepting PANAS Negative, Global Life Satisfaction and WHO-5) on self deception. Overall the performance of WEMWBS against the BIDR was therefore good.

Table 37: Correlation between BIDR and scales used in questionnaire pack

	Impression Management	Self-Deception
WEMWBS	0.18*	0.35**
Affectometer 2	-0.25**	0.55**
Satisfaction with Life Scale	0.34**	0.40**
WHO-5	-0.39**	-0.20
PANAS-PA	0.02	0.50**
PANAS-NA	0.03	-0.16
Global Life Satisfaction	0.26*	0.13

*significant at 0.05 level

** significant at 0.01 level

6.3.9 Correlation between Affectometer 2 and WEMWBS

The correlation between Affectometer 2 and WEMWBS was tested on a convenience sample of 72 adults. It was hypothesised that there should be a moderate to high correlation between the two scales, given that the positive and negative scales of Affectometer 2 are only moderately highly correlated and that adjective scales had a different mean from sentence scales and that WEMWBS comprises only positive statements about positive attributes. The two scales had a moderate to high correlation of 0.81 ($p < 0.01$) (Pearson's correlation coefficient).

6.4 Discussion of WEMWBS findings

6.4.1 Descriptive information

Scores on the Warwick-Edinburgh Mental Well-Being Scale followed a Normal distribution suggesting neither floor nor ceiling effects and thus suitability for use in monitoring positive mental health at a population level. Factor analysis of the WEMWBS suggested a single underlying construct indicating that it is appropriate to summarise all items into a single score. This result stands in contrast to the results on the Affectometer 2 which were more equivocal. This difference is likely to reflect the positive nature of the scale.

Like Affectometer 2, WEMWBS is capable of distinguishing between different population groups, detecting significant differences in mean scores between different universities and different subject groups. Differences across age groups were detected in HEPS data for Affectometer 2 but not in the student sample. The fact that no differences were detected across age groups with WEMWBS with student samples may therefore be a feature of the relatively narrow age band on which WEMWBS has been tested and needs investigating in population samples. The lack of statistically significant difference between men and women also needs further investigation: it would not appear that gender differences are detectable with WEMWBS, although this may be due to the low proportion of men who took part.

NOT FOR WIDER CIRCULATION OR REPRODUCTION

This finding needs validating in a population sample as other measures of mental health show strong gender differences.

Although the response rate was lower for this phase of testing on WEMWBS than for the equivalent phase for Affectometer 2, the required number of respondents for this phase was an overestimate of the actual number needed. This is because the sample size calculation was based on the number of responses needed to carry out the factor analysis, which is based on the number of items in the scale. Because WEMWBS is a much shorter scale than Affectometer 2, the actual number of responses required in this phase was smaller than first calculated and the number of responses received was sufficient for the statistical analyses used in the research. However, it has not been possible to assess whether there was any difference in the people who completed the scale and those who did not (response bias).

6.4.2 Internal consistency

Although reducing the length of Affectometer 2 had the expected effect of reducing the very high internal consistency score to a more appropriate level, the alpha coefficient for WEMWBS remains relatively high at 0.89. There therefore appears to be potential to reduce the scale further to around nine items.

6.4.3 Construct validity

WEMWBS showed a similar pattern of construct validity to Affectometer 2 with most of the scales tested, with a slightly higher correlation with the Satisfaction with Life Scale and a slightly lower correlation with the single-item Global Life Satisfaction Scale. The reason for these differences is not clear but may reflect differences in the concepts captured by each of the life satisfaction scales. A higher correlation was observed between WEMWBS and the positive scale of PANAS, reflecting the fact that, unlike Affectometer 2, WEMWBS only includes positive items. Correlations with two of the scales that were included because they measure concepts related to, but different from, positive mental health (the Emotional Intelligence Scale and EQ-5D) were similar when testing both Affectometer 2 and WEMWBS. Correlations between Affectometer 2 scores and WHO-5 and the Scale of Psychological Well-Being were higher than correlations between WEMWBS and the same two scales.

6.4.4 Test-retest reliability

WEMWBS showed a similarly high one week test-retest reliability to that shown by Affectometer 2, suggesting that test scores are not unduly influenced by current mental states.

6.4.5 Response bias

WEMWBS is less susceptible to response bias, as measured using the BIDR, than Affectometer 2, showing the lowest impression management bias of all the scales tested, with the exception of the two PANAS scales. Unlike Affectometer 2, WEMWBS showed a positive correlation with the impression management sub-scale suggesting that respondents with high WEMWBS scores were slightly more likely to attempt to give a favourable impression of themselves than people with low WEMWBS scores. However the correlation was low (0.18) and lower than those correlations observed with two of the five other scales. Like Affectometer 2, WEMWBS showed a positive correlation with the self-deception sub-scale, but the correlation was lower (0.35 for WEMWBS, 0.55 for Affectometer 2) and also lower

than those observed for other comparable scales. This suggests that on both the two BIDR sub-scales, WEMWBS is less susceptible to response bias than other comparable measures.

6.4.6 Correlation between Affectometer 2 and WEMWBS

Although there was a high, statistically-significant correlation between Affectometer 2 and WEMWBS (0.81), this was lower than anticipated, given that WEMWBS is derived from Affectometer 2. This may be the result of removing the negative items from the scale and turning it into an exclusively positive scale. This hypothesis is supported by the fact that positive and negative items in Affectometer 2 do not show a perfect correlation. However, the strength of the correlation, which is higher than any of the correlations observed between either Affectometer 2 and WEMWBS and the other scales to which they were compared, suggests that the revised scale is tapping into a very similar concept to that measured by Affectometer 2, and can therefore be regarded as a suitable alternative measure of positive mental health.

7 Face validity of WEMWBS

7.1 Aims

The focus groups had six aims:

- to test what participants thought WEMWBS was designed to measure
- to test how easy participants found it to complete WEMWBS and in particular, whether they found the statement or word components of the scale easier
- to discuss participants' understanding of the terms 'mental health', 'mental well-being' and 'positive mental health'
- to discuss whether people thought WEMWBS was a measure of positive mental health
- to identify any items that people associated with mental well-being which were not covered by the scale
- to identify items that people did not understand, or did not think were relevant.

7.2 Method

Once the psychometric properties of WEMWBS had been tested, two extra focus groups were held, one in Edinburgh, Scotland and one in Leamington Spa, England. Participants were purposively selected to represent a range of different ages, sexes, socio-economic backgrounds, and included users of mental health services. Interviews followed an amended version of the interview schedule used for the Phase 2 focus groups.

7.3 Results

Seven participants took part in the two focus groups (Table 38).

Table 38: Summary of participants

Group	Location	Number of attendees
General Group (18-64)	Leamington Spa, England	4
General Group (18-64)	Edinburgh, Scotland	3

7.3.1 What the tool is measuring

There was a consensus across the two groups that WEMWBS was about mental health, how people felt on a day-to-day basis and about their current circumstances.

'Well I though it was like your mental health, like...stuff like that.'
(General Scottish Group 18+)

'Because it's about how you feel'
(General Leamington Spa Group 18+)

NOT FOR WIDER CIRCULATION OR REPRODUCTION

7.3.2 *Ease of completion*

Both focus groups agreed that WEMWBS was straightforward and easy to complete, and it was commented on by the Scottish group that they liked it because the scale items were honest questions:

'It's just simple questions with simple answers'

(General Scottish Group 18+)

'That's one of the things I liked about it. It's not full of jargon and it's easy to understand'

(General Scottish Group 18+)

'There's no hidden agenda and I thought it was just...I was just having to think clearly about have I or have I not do you know what I mean.'

(General Scottish Group 18+)

Participants commented that they gave a gut reaction when completing WEMWBS; and that they would have found it more difficult to complete if they had been given prior notice.

7.3.3 *How the scale might be used*

Participants came up with a number of potential uses for the scale. It was suggested that WEMWBS could be used by the National Health Service to assess the need for provision in the future, and to assess the health of the nation:

'I'm not being cynical, but is the National Health Service trying to find out what sort of prognosis there is for the future, how many patients are going to find it tough for various reasons.'

(General Leamington Spa Group 18+)

'Well, it's a possibility if they need to know in advance more of this and what's coming along in the future for all the country if people are getting older or whatever it may be, or wherever the groups are, they need to make provisions of some sort no doubt.'

(General Leamington Spa Group 18+)

7.3.4 *Limitations on scale use*

Although participants were generally at ease completing the scale, they suggested that there were certain circumstances in which they felt it should not be used:

'No, it still would depend who had...who gives you the questions no matter what. If I went to my GP's surgery and I'd just given birth and they asked me those questions, I would think 'Well, they're treating me for post-natal depression'.'

(General Scottish Group 18+)

Participants also said that they felt more comfortable completing the scale with the researchers because they did not know them.

NOT FOR WIDER CIRCULATION OR REPRODUCTION

7.3.5 Scale time-scale

Both focus groups were interested in why the scale instructions asked them to think back over the previous two weeks. Participants admitted that they were filling in the scale 'generally', i.e. they were responding to how they have felt in the last couple of years rather than focussing on the previous two weeks. However, participants noted that their responses could be different depending on their circumstances and situation at that particular time:

'So, I was thinking if it had been two weeks prior to that it might have been slightly different and so I just kind of went with that.'

(General Leamington Spa Group 18+)

'Yes, we're just filling it in about life in general as in...compared to two weeks.'

(General Scottish Group 18+)

Despite discussions about what time-scale they felt would be suitable, it was agreed that changing from two weeks would become confusing, and that a two-week time period enabled participants to record a snapshot of their life.

7.3.6 Definitions of mental health

7.3.6.1 Mental health

During the focus groups, participants were asked how they defined mental health. The focus group moderator tested responses to different words that are associated with mental state, including 'mental health', 'mental illness', 'well-being' and 'emotional health'.

Mental health was viewed by both groups as a person's state of mind, and how a person functions. Focus groups also stated that mental health is present in everyone, and situations and circumstances can cause a person to experience mental ill health. One participant commented that WEMWBS was inclusive as everyone would be able to complete the scale and would feature within the range of scores:

'Well I think the way the questions are phrased it's very...I mean mental health applies to *everybody*, so I suppose...you know...everybody would be on the range somewhere.'

(General Leamington Spa Group 18+)

'So, mental health...I think is your state at that time and how you respond to *everything* in your life.'

(General Scottish Group 18+)

'Yes, well everybody's got their own mental health, right? If it was something like an illness, then it's how you cope with than and then what pushes you over the edge, or whether you actually just stay in the exact same state of mental health that you were in before that incident happened.'

(General Scottish Group 18+)

NOT FOR WIDER CIRCULATION OR REPRODUCTION

7.3.6.2 *Mental health and physical health*

Focus groups also commented that there is an interaction between and mental and physical health:

‘Well I mean...you can link the two together – in physical disability for example, if you’ve got mobility problems. That could bring on a sense of feeling low or depressed, which isn’t generally regarded as mental illness, which could lead to a chaotic lifestyle or making wrong decisions in different situations.’

(General Scottish Group 18+)

‘Well my sister has got cancer and it’s definitely affecting her mentally.’

(General Scottish Group 18+)

7.3.6.3 *Mental illness and mental well-being*

Participants were asked whether they thought that an individual could experience mental illness yet enjoy mental well-being. There appeared to be a consensus that if the individual’s difficulties were managed properly then they could experience mental well-being. In addition, participants commented that those experiencing mental illness do not experience difficulties throughout their illness:

‘If you’ve got a mental illness and it’s being managed, you can still have enjoyment and pleasure out of life.’

(General Scottish Group 18+)

‘Mental illness is in a sense a label, but it doesn’t necessarily define how you are all the time.’

(General Scottish Group 18+)

Participants commented that, while an individual might perceive that they are coping with their situation, the views of those around them, either friends/family or health professionals, might be very different:

‘Well you can be mentally ill but yourself think that you’re coping really well and that you’re you know...you’re OK and yet other people might think you’re falling to pieces. I mean you might not see yourself as being like that, so therefore you think you’re OK and you feel well in yourself, but a doctor or an outside person might think...you know...you’re not well at all, but you might not be able to see that yourself because you think you’re OK.’

(General Scottish Group 18+)

7.4 Key Conclusions

Participants attending the focus groups found the scale unambiguous and easy to complete. Unlike the focus groups carried out to discuss Affectometer 2, no participants made any comments about how the scale could be modified or improved, suggesting that WEMWBS may be easier to complete than Affectometer 2.

NOT FOR WIDER CIRCULATION OR REPRODUCTION

In this phase of the research, participants described the scale as measuring mental health. Later discussions in the focus group showed that participants in this phase were more likely to conceptualise mental health as an emotional state experienced by everyone which was distinct from mental illness. Participants appeared to have a holistic view of mental health, which encompassed physical and mental states and which was affected by ill-health. This broad conceptualisation of mental health fits more closely with the definition of positive mental health used by this research and, more widely, by NHS Health Scotland.

The focus groups also shed some light on whether WEMWBS is measuring state or trait. Participants queried the use of the two-week timeframe which they were asked to use as a reference point when completing the scale, but appeared to recognise that this was a way of encouraging them to reply in a way which reflected a more general trait. Nevertheless, participants also recognised that their responses could change over time, and as their circumstances changed.

One important issue raised by the focus groups which did not come out as clearly in the first set of focus groups was the issue of when and how the scale should be used. Some participants clearly expressed concern about it being used as a diagnostic tool by health professionals. This view appeared to be based on mistrust of health professionals rather than a reflection on WEMWBS, as they expressed similar concerns about other screening tools, such as the Edinburgh Post-natal Depression Scale. However, the focus groups seemed to feel that it would be a good tool to use at a population level, where their responses would be anonymous.

Overall, these focus groups show that WEMWBS has a high degree of acceptability in the general population and is easy to understand and complete. It appears to fit in with participants' perceptions of what mental health is, which the group defined as a general state which is distinct from mental illness. Although the number of participants taking part in this phase of research was small, the results are consistent with WEMWBS having good face validity in the general population. Further studies with greater numbers would be welcome. The focus groups appear to support the use of WEMWBS as a tool to measure population health and came up with a number of plausible reasons about how it might be used at this level.

8 General discussion of findings

8.1 Underlying purpose of Affectometer 2 and WEMWBS

A key element in the validation of new scales is to describe what the scale is intended to measure and scale validation techniques generally work best when scales are intended to measure a well defined concept and can be tested against an objective measure (such as a clinical diagnosis of disease). Whilst there remains some discussion and debate about the nature of positive mental health, this aspect of validation of the Affectometer 2 and WEMWBS scales must remain an inexact science. In the case of positive mental health, there are no objective measures or even a clearly and widely accepted definition. The best that can be offered at this stage is a combination of quantitative data assessing the performance of these scales against other scales using standard validation techniques with qualitative feedback from potential users of the scale and experts in the field.

In the case of both Affectometer 2, and the new scale, WEMWBS, the conceptual underpinnings accord with accepted definitions of 'positive mental health' reflecting current concerns of the policy community and expert opinion about what feelings or behaviours contribute towards this state. But to some extent they also contravene accepted definitions. For example, one of the strengths of both instruments, from a pragmatic point of view – that they cover both hedonic (subjective experience of happiness and life satisfaction) and eudaimonic (psychological functioning, relationships with others) components might be considered a weakness by the purist in positive psychology or subjective well-being research.

The formal and statistical validation processes used in this validation study attempt, as far as is possible, to assess whether the scale captures a single concept, which we define as 'positive mental health' and whether it is measuring this concept, when compared to other similar tests.

Affectometer 2, like many mental health scales covers both positive and negative concepts of mental health in positive and negative statements. The score of negative items is subtracted from positive items. The factor analysis on the HEPs dataset, whilst generally supporting a single underlying factor, failed to provide unequivocal results to show that this was an acceptable method. Those who believe that positive and negative mental health are distinct concepts might interpret the results as suggesting that the scale scoring method was inappropriate. WEMWBS with its entirely positive statements of only positive concepts gets away from this problem and in this respect is unlike any other scale of mental health. The fact that the factor analysis on WEMWBS pointed clearly to a single concept, validates this approach.

8.2 Is the scale measuring 'positive mental health'?

This research has used several processes to address this question. First, focus groups were used to assess what members of the general public thought Affectometer 2 and WEMWBS were measuring. This was not an easy question for participants to answer. Just as the academic and policy communities have not yet come up with a clear definition of 'positive mental health', the general public struggles to distinguish between mental health, positive mental health, well-being and mental

illness. However, the focus groups did suggest that at least certain sections of the population understand, and are able to articulate, a mental state described by several participants as 'how you feel in yourself' which they felt Affectometer 2 expressed. Such views were expressed more strongly and clearly by the small number of participants who took part in the focus groups held to discuss WEMWBS, suggesting that the new scale may have higher face validity than Affectometer 2. The focus on positive items is likely to play a part in this. Although Affectometer 2 contains a balance of positive and negative items, focus groups consistently said that they thought the scale was predominantly negative.

The second method used to assess whether the Affectometer 2 and WEMWBS measure positive mental health was to test their construct validity as compared to scales measuring general health, life satisfaction, positive and negative affect or mood, happiness and depression. Although this is a key part of scale validation processes it is inherently problematic when used on scales for which the expected overlap is not known, or which can only be roughly estimated. For this reason, expert opinion was used to determine the degree to which scales should correlate with each other but precise correlations could not be developed. Correlations were generally in line with what was expected, giving a broad indication that the scale is measuring a concept that overlaps with several other related concepts, but has a distinctive component that is not covered by other scales.

8.3 Do Affectometer 2 and WEMWBS measure the same concept?

The aim of this research was to validate Affectometer 2 and, if necessary, develop a shorter/new version of the scale. As part of this process, it has also been necessary to establish whether or not Affectometer 2 and WEMWBS measure the same concept. This has been done in two ways: first, as a part of the process of developing the new scale, the expert panel agreed what were the important components of positive mental health that were covered by Affectometer 2 and incorporated these in the new scale. With the exception of making WEMWBS an exclusively positive scale, with the possible exception of 'purpose in life' and a different approach to capturing 'autonomy', the two scales cover the same components. Secondly the extent of correlation between Affectometer 2 and WEMWBS scores were measured on a convenience sample. Results indicated a high, but not perfect correlation between the two, suggesting a high degree of overlap. The lack of a perfect correlation may be due to two factors: the lack of negative items and concepts in WEMWBS may have affected scoring patterns, or the two scales may be measuring slightly different concepts. For practical purposes, however, the two scales appear to be closely aligned.

8.4 Scale completion and use

This research provides important practical information about how easy the scale is to complete. Both sets of focus groups found the scales generally easy to fill in. Older people had more difficulties than younger people, but this is true of all scales. Response rates to both Affectometer 2 and WEMWBS were also high in the student samples and compared favourably to response rates for the GHQ-12 in the HEPS dataset.

NOT FOR WIDER CIRCULATION OR REPRODUCTION

Feedback from the focus groups allowed modifications to be made to Affectometer 2, such as clarifying the response categories to draw a sharper distinction between one level of response and the next, and only including sentences in the scale. Although few participants commented on the length of Affectometer 2, the practical advantages of a shortened scale, particularly if it is to be included as part of a battery of questionnaires, are obvious and could be realised without compromising on the overall performance of the scale.

The research points to some potential caveats about the use of the scale. Although participants in both sets of focus groups were happy to complete the scale and felt it would be a good tool to measure people's positive mental health at a population level or in a group setting, some participants expressed reservations about how the tool might be used at an individual level. In particular, concerns were raised about the tool being used by health professionals to gather information about an individual's mental health state (people who expressed this view had similar reservations about other routinely-used screening tools, such as the Edinburgh Post-natal Depression Scale). This view appears to reflect a broader mistrust of health professionals rather than specific concerns about either Affectometer 2 or WEMWBS.

Currently, WEMWBS scores are presented as means with confidence intervals and in graphical format. Cut off points to indicate low or high levels of positive mental health cannot be defined on the basis of the data presented in this report. Research in progress on a large population sample in Scotland, where WEMWBS has been completed along with the GHQ-12, may enable cut off points to be defined at the same time as providing population norms.

8.5 Strengths and weaknesses of this research

This research used a variety of techniques (qualitative and quantitative) to test the validity of Affectometer 2 and to develop and test WEMWBS. The use of focus groups in scale validation is not yet routine. In this research, members of the general population were involved in conceptual discussions about what the scale was measuring and how it could be improved to reflect the feelings and emotions that make up positive mental health as well as practical issues about how the scale could be made easier to use. This resulted in a number of modifications, clearer response categories and the removal or rewording of certain items. The focus groups also provided valuable insights into some of the terminological problems in this area. Participants generally associated the term 'mental health' with mental illnesses. The term 'well-being' appeared to be largely associated with physical, rather than mental health. The title of the revised scale, which incorporates both 'mental' and 'well-being' attempts to bridge the gap between the two concepts of 'mental health' and 'well-being'.

Several of the participants in the general population focus groups and all of those in the mental health service user group had experienced mental illness. The general men's group were recruited from a project set up to help men to cope with mental health problems, and several women in the socio-economic groups C2, D, E women's group had experienced episodes of clinical depression. This analysis suggests that participants who had overcome, or were living with, mental health problems were more likely to be familiar with concepts of positive mental health or well-being than those participants who did not have direct experience of mental

NOT FOR WIDER CIRCULATION OR REPRODUCTION

illness. The results of these focus groups may suggest a greater ease and familiarity with mental well-being concepts than is the case in the general population.

Because of resource constraints, the two main rounds of testing were carried out with student samples. Students are younger, better educated and likely to come from more affluent families than the general adult population. Although it is common practice to use student populations for scale development and testing because of the costs of recruiting large samples in the general population, this may have compromised the generalisability of our findings. Attempts were made to combat this potential problem using as broad an age-range as possible (post-graduate and mature students as well as undergraduates) and by involving students from different subjects in different countries. Comparison of scores derived on Affectometer 2 in the student and general population samples using data collected in HEPS do not suggest that this was a significant problem. However, testing WEMWBS on a larger representative sample of the population is recommended. Such a study would permit assessment of the scale's capacity to discriminate between subgroups on the basis of factors such as age, sex, marital status, and social class. In the first instance, WEMWBS is to be included in the September 2006 wave of HEPS to enable these associations to be tested (report expected Summer 2007). The results will also be compared to the earlier analysis of HEPS data.

In this project we have not been able to test the sensitivity to change of either the Affectometer 2 or the WEMWBS. Such research is needed before measures can be recommended for use in evaluation studies or intervention trials where the aim is to measure improvements in health resulting from the interventions. Our literature review revealed a number of studies in which the sensitivity to change had been demonstrated for the Affectometer 2. Whilst this augurs well in terms of WEMWBS, sensitivity to change cannot be guaranteed. Scales with smaller numbers of items are, other things being equal, less sensitive to change than those with more items, because there is more opportunity for respondent scores to change. Until the latter research has been undertaken for WEMWBS, it would be better to use the Affectometer 2 in such studies.

Alternative techniques are available to test the psychometric properties of ordinal scales, such as Affectometer 2 and WEMWBS, that have not been employed in this research. Techniques such as Rasch analysis – an alternative method for evaluating the psychometric properties of scales (Bond & Fox, 2001), could be used on data collected from a population sample such as HEPS to further explore WEMWBS's psychometric properties and to determine whether the tool can be used at an individual as well as a population level. Such analyses combined with further analysis of internal consistency in a general population sample may point to opportunities to reduce the number of items in the scale to nine or ten. Such analyses of the data from WEMWBS in the September 2006 wave of HEPS is planned for early 2007.

9 Overall conclusions

9.1 Suitability of WEMWBS as a measure of positive mental health at a population level

Both Affectometer 2 and WEMWBS have been rigorously tested against current standards for scale validation. Although both scales have good face validity and reflect current concepts of positive mental health, WEMWBS appears to be a more suitable measure than Affectometer 2 because it is shorter, making it more practical for population survey, and is less susceptible to response bias. Some of the practical difficulties of completing Affectometer 2 that focus group participants identified have also been eliminated, in particular difficulties distinguishing between scale response categories and problems interpreting the 20 word items in Affectometer 2.

WEMWBS was well received by focus group participants who were also positive about the suggestion that this tool may in the future be used to measure positive mental health of adults at a population level.

WEMWBS showed appropriate levels of construct validity when compared with other scales. Lack of conceptual clarity about the relationship between concepts such as mental health, mental illness, well-being, positive mental health, life satisfaction etc makes it difficult to assess construct validity in a robust way. However, this research showed that WEMWBS appears to relate plausibly to other scales.

WEMWBS also showed an appropriate stability over time: although the scale needs to be able to capture changes in people's positive mental health over time, it should be relatively stable over a short period of time. This research suggests that people's scores at week one are strongly correlated with their score at week 2, indicating that the scale is not susceptible to short-term changes in mood.

Although there is a range of well-established scales that have been used to measure concepts related to positive mental health, some of which were tested in this research, WEMWBS has been developed specifically to measure positive mental health and should therefore be considered for use in this context. However, it is less likely to be suitable for measuring related concepts such as emotional intelligence or life satisfaction, for which specific scales are available.

9.2 Potential future uses of WEMWBS

This research suggests that WEMWBS is likely to be well-received by a wide range of different adult population groups and is easy and quick to complete. It is therefore likely to be in demand as an evaluation tool to measure the impact of policies, programmes or interventions that are aiming to influence positive mental health. However, further research needs to be carried out to identify whether the tool is capable of measuring changes in positive mental health before and after interventions. In the meanwhile organisations wanting to evaluate the impact of their policies, programmes or interventions on positive mental health may want to consider using Affectometer 2.

9.3 Suitability of WEMWBS for use at an individual level

The remit of this research only covers the use of Affectometer 2 and WEMWBS at a population level. The developers of Affectometer 2 state that the scale may be used at an individual therapeutic level, with the participant completing the scale, working out their own score and then using this as a discussion tool with their counsellor or course leader. However, no examples of the scale being used in this way were identified in our literature review.

Discussion with participants in focus groups suggest that members of the public may find this use of the scale uncomfortable and intrusive particularly if administered by someone in a position of power with whom a trust relationship has not been established.

9.4 Affectometer 2 or WEMWBS?

This research indicates that Affectometer 2 performed well against most accepted criteria. Its main drawback is its length and its apparent susceptibility to response bias. As a shorter measure, which is less prone to response bias than either Affectometer 2 or other comparable measures, WEMWBS may be a preferable scale to use. This recommendation however must remain tentative until the measure has been tested in population samples and sensitivity to change demonstrated.

References

Bond, T.G. & C.M. Fox (2001) *Applying the Rasch Model: fundamental measurement in the human sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates

Boyle, G.J. (1991) 'Does item homogeneity indicate internal consistency or item redundancy in psychometric scales?' *Personality and Individual Differences* 12 (3): 291-294

Cohen, J. (1988) *Statistical power analysis for the behavioural sciences* (2nd Ed) Hillsdale, NJ: Lawrence Erlbaum Associates

Commonwealth Department of Health and Aged Care (2002) *Promotion, prevention and early intervention for mental Health - a monograph* Canberra: Commonwealth Department of Health and Aged Care

Dennerstein, L., A.M. Smith, & C. Morse (1994) 'Psychological well-being, mid-life and the menopause'. *Maturitas* 20 (1): 1-11

Dennerstein, L., P. Lehert, E. Dudley, & J Guthrie (2001) 'Factors contributing to positive mood during the menopausal transition'. *Journal of Nervous and Mental Disease* 189 (2): 84-89

Dennerstein, L., P. Lehert, & J. Guthrie (2002a) 'The effects of the menopausal transition and biopsychosocial factors on well-being'. *Archives of Women's Mental Health* 5 (1): 15-22

Dennerstein, L., J. Randolph, J. Taffe, E. Dudley & H. Burger (2002b) Hormones, moods, sexuality and the menopause transition. *Fertility and Sterility* 77 (S4): 42-8

Department for Education and Skills. (2004) *Every Child Matters: next steps* London: TSO

Department of Health (2004) *National Service Framework for Children and Young People and maternity services* London: TSO.

European Commission (2005) *Promoting the mental health of the population. Towards a strategy on mental health for the EU* COM (2005) 484 Brussels: European Commission

Flett, R., B. Harcourt, & F. Alpass (1994) 'Psychosocial aspects of chronic lower leg ulceration in the elderly'. *Western Journal of Nursing Research*, 16 (2): 183-92

Goldberg, D.P., & P. Williams (1988) *A user's guide to the General Health Questionnaire*. Windsor, UK: NFER-Nelson

Hemphill, J.F. (2003) 'Interpreting the magnitudes of correlation coefficients' *American Psychologist* 58(1): 78-9

NOT FOR WIDER CIRCULATION OR REPRODUCTION

- Hossack, R.C. (1997) Salutogenic and pathogenic orientations to life: attachment, personality, sense of coherence and well-being in late adolescence: a structural equation model. *Dissertation Abstracts International: Section B: The Sciences and Engineering* 58 (4B): 2153
- Kammann, R. & R. Flett (1983a) Affectometer 2: a scale to measure current level of general happiness. *Australian Journal of Psychology* 35(2): 259-265
- Kammann R & R. Flett (1983b) Sourcebook for Measuring Well-Being with Affectometer 2. Dunedin: Why Not? Foundation
- Kimweli, D. (2002) The relationship of gender-role development, college level and subjective well-being among Appalachians. *Dissertation Abstracts International: Section B: The Sciences and Engineering*. 63 (1B): 531
- Lo, J., R. Zemke (1997) The relationship between affective experiences during daily occupations and subjective well-being measures: a pilot study. *Occupational Therapy in Mental Health* 13 (3): 1-21
- Lohr, K.N., N.K. Aaronson, J. Alonso, M. A. Burnan, D.L. Patrick, E.B. Perrin, & J.S. Roberts (1996) Evaluating quality-of-life and health status instruments: development of scientific review criteria. *Clinical Therapeutics* 18 (5): 979-992
- Nunnally, J.P. (1978) *Sychometric Theory* New York: McGraw-Hill
- O'Byrne, M.C. (1997) Paternal responsibility, involvement with children, and sense of well-being. *Dissertation Abstracts International: Section B: The Sciences and Engineering*. 57 (10B): 6585
- Peters, E., M.P. Wissing & W.F. du Plessis (2002) Implementation of EMDR with cancer patients. *Health SA Gesondheid*, 7 (2): 100-109
- Raeburn, J.M., J.M. Atkinson, J.M. Dubignon, J. Fitzpatrick, M. McPherson & G.S. Elkind (1994) 'Superhealth Basic: an evaluation of a low-cost community-based lifestyle change programme'. *Psychology & Health* 9(5): 383-395
- Rand, C., D.L. Graham & E.I. Rawlings (1982) 'Psychological health and factors the court seeks to control in lesbian mother custody trials'. *Journal of Homosexuality*. 8 (1): 27-39
- Robinson, J.P., P. Shaver & L. Wrightsman (1991) *Measure of personality and social psychological attitudes* San Diego, California: Academic Press
- Ryan, R.M. & E.I. Deci (2001) 'On happiness and human potential: a review of research on hedonic and eudaimonic well-being'. *Ann Rev Psych* 52: 141-166.
- Salovey, P. & J.D. Mayer (1990) 'Emotional intelligence' *Imagination, Cognition and Personality* 9: 185-211

NOT FOR WIDER CIRCULATION OR REPRODUCTION

Schmitt, N. (1996) 'Uses and abuses of coefficient alpha'. *Psychological Assessment* 8 (4): 350-353

Scottish Executive (2003) *National programme for improving mental health and well-being: action plan 2003-6* Edinburgh: Scottish Executive.

Singleton, H.L.G. (ed.) (2003) *Better or worse: a longitudinal study of the mental health of adults living in private households in Great Britain* London: The Stationery Office

Speight, J, C. McMillan, M. Barrington, & C. Victor (forthcoming) *Review of scales of mental health* NHS Health Scotland

Stein, I.L. (1996) *Inter-institutional relocation of aged people*. PhD, University of Wollongong

Stewart-Brown, S. (2002) Measuring the parts most measures do not reach: a necessity for evaluation in mental health promotion. *Journal of Mental Health Promotion* 1(2): 4-8

U.S. Department of Health and Human Services (2001) *Mental Health: Culture, Race and Ethnicity - A Supplement to Mental Health: a report of the Surgeon General* U.S. Department of Health and Human Services, Substance Abuse & Mental Health Services Administration.

World Health Organisation (2001). *Strengthening mental health promotion*, in *Fact Sheet, No.220* Geneva: WHO.

Appendices

Appendix 1: Affectometer 2

1 – STATEMENTS

Look at each statement and circle the number that best fits how often you've felt that way in the last 2 weeks.

Example question

- One of the statements is 'I have energy to spare'. If you haven't felt as if you've had any energy to spare at all in the last 2 weeks, please tick 'not at all'. If you've had energy to spare all of the time in the last 2 weeks, tick 'all of the time'.

	Not at all	Occasionally	Some of the time	Often	All of the time
I have energy to spare	0	1	2	3	✓4

STATEMENTS	Not at all	Occasionally	Some of the time	Often	All of the time
My life is on the right track	0	1	2	3	4
I have been left alone when I don't want to be	0	1	2	3	4
I feel I can do whatever I want to	0	1	2	3	4
I have been thinking clearly and creatively	0	1	2	3	4
Like a failure	0	1	2	3	4
Nothing seems very much fun any more	0	1	2	3	4
I like myself	0	1	2	3	4
I can't be bothered to do anything	0	1	2	3	4
Close to people around me	0	1	2	3	4
As though the best years of my life are over	0	1	2	3	4
My future looks good	0	1	2	3	4
I have lost interest in other people & don't care about them	0	1	2	3	4
I have energy to spare	0	1	2	3	4
I smile and laugh a lot	0	1	2	3	4
I wish I could change some part of my life	0	1	2	3	4
My thoughts go around in useless circles	0	1	2	3	4
I can handle any problems that come up	0	1	2	3	4
My life seems stuck in a rut	0	1	2	3	4
I feel loved and trusted	0	1	2	3	4
I feel there must be something wrong with me	0	1	2	3	4

NOT FOR WIDER CIRCULATION OR REPRODUCTION

2 - WORDS

All you need to do is to look at each word and circle the number that best fits how often you've felt that way in the last 2 weeks.

Example question

- One of the words is 'lonely'. If you haven't felt lonely at all in the last 2 weeks, please tick 'not at all'. If you've felt lonely all of the time in the last 2 weeks, tick 'all of the time'.

	Not at all	Occasionally	Some of the time	Often	All of the time
Lonely	0	1	2	3	✓4

Words	Not at all	Occasionally	Some of the time	Often	All of the time
Satisfied	0	1	2	3	4
Lonely	0	1	2	3	4
Good natured	0	1	2	3	4
Clear headed	0	1	2	3	4
Helpless	0	1	2	3	4
Impatient	0	1	2	3	4
Useful	0	1	2	3	4
Depressed	0	1	2	3	4
Relaxed about things	0	1	2	3	4
Hopeless	0	1	2	3	4
Optimistic	0	1	2	3	4
Withdrawn	0	1	2	3	4
Enthusiastic	0	1	2	3	4
Interested in other people	0	1	2	3	4
Discontented	0	1	2	3	4
Confused	0	1	2	3	4
Confident	0	1	2	3	4
Tense	0	1	2	3	4
Understood	0	1	2	3	4
Insignificant	0	1	2	3	4

Permission to use this scale should be sought from the author, Dr. Ross Flett, School of Psychology, Massey University, Private Bag 11-222, Palmerston North, New Zealand. Email. r.a.flett@massey.ac.uk.

Appendix 2: Analysis of HEPS dataset

This is an unpublished draft of a paper which has been submitted for publication. A revised paper has been accepted in the journal Quality of Life Research. Not for reproduction or quotation without the authors' permission.

The Affectometer 2: a valid measure of positive mental health in UK populations

Ruth Tennant¹, Stephen Joseph², Sarah Stewart-Brown¹

¹ Warwick Medical School, University of Warwick, Coventry, CV4 7AL, UK. (Email: r.tennant@warwick.ac.uk)

² Department of Sociology and Social Policy, University of Nottingham (Email: Stephen.Joseph@nottingham.ac.uk)

Address for correspondence: Sarah Stewart-Brown, Warwick Medical School, University of Warwick, Coventry, CV4 7AL, UK

Phone: +44 24 7657 4510

Fax: +44 24 7657 4879

Email: s.stewart-brown@warwick.ac.uk

Introduction

There is increasing international interest in the concept of positive mental health and its contribution to all aspects of human life. The World Health Organisationⁱ has declared positive mental health to be the ‘foundation for wellbeing and effective functioning for both the individual and the community’ and defined it as a state ‘which allows individuals to realise their abilities, cope with the normal stresses of life, work productively and fruitfully, and make a contribution to their community’. The capacity for mutually satisfying and enduring relationships is another important aspect of positive mental healthⁱⁱ.

The term positive mental health is often used, in both policy and academic literature, interchangeably with the term mental wellbeing, and mental wellbeing may be represented simply as wellbeing. Mental wellbeing is a complex construct, covering both experience and functioning with two distinct perspectives:- the hedonic perspective which focuses on the subjective experience of happiness and life satisfaction and the eudaimonic perspective favoured by the positive psychologists, focusing on psychological functioning and self realisationⁱⁱⁱ. These perspectives, which have informed distinct bodies of research in positive mental health, are less obvious in the literature relating to poor mental health where items measuring affect (feeling happy/sad) are often combined with items measuring positive functioning (playing a useful part in things, making decisions)^{iv} in the same scales, suggesting that poor mental health at least is accepted as involving limitations in both eudaimonic and hedonic wellbeing.

Some researchers with an interest in positive mental health have proposed that this construct is independent of mental illness. They observe that people with mental illness have varying levels of both subjective wellbeing and psychological functioning,

NOT FOR WIDER CIRCULATION OR REPRODUCTION

and that sizeable proportions of the population who do not have mental illness lack wellbeing^{v.vi}. Whilst few studies have shown positive and negative aspects of mental health to be entirely independent, they may represent separate but correlated dimensions of health⁵. Whilst discussion continues about the precise nature of positive mental health, policy makers are being driven, in part by the priorities of the general public^{vii}, to embrace the promotion of positive mental health and wellbeing^{viii,ix} in addition to preventing mental illness, and therefore have a need to know whether their policies are having the desired impact. As a result there is increasing interest in the measurement of positive mental health. Historically, mental health measurement tools have focused on psychiatric morbidity, dividing the population into those who meet the criteria for diagnosis of mental illness and those who are 'normal'. These tools are not well-suited to measuring mental health at a population level as they show 'ceiling effects' with most people scoring the maximum possible score. They are therefore unable to distinguish average from good mental health^x.

A recent review of measures covering aspects of positive mental health identified one – Affectometer 2 - that appeared to relate well to the World Health Organisation's definition of mental health¹. This measure^{xi}, developed in New Zealand in 1983, covers a range of aspects of positive mental health including subjective well being, psychological functioning and relationships. It includes many items not covered in scales of negative mental health (feeling loved and trusted, thinking clearly and creatively, problem solving) and, also unlike the latter, has equal numbers of items relating to the positive as negative end of the spectrum of health. Its coverage, balance of positive and negative items, and intuitive appeal to practitioners and policy makers in the UK makes it a promising candidate as a tool for monitoring positive

NOT FOR WIDER CIRCULATION OR REPRODUCTION

mental health and suggests the need for an assessment of its validity and reliability in the UK population.

The developers of Affectometer 2 say that it was developed to measure well-being, which they defined as the overall balance between good and bad feelings or emotions. It has also been described as a 'general tool to measure happiness'. The underlying theoretical principle of the scale is that an individual's mental health status is determined by the degree to which positive feelings, attitudes and beliefs outweigh negative ones. It was intended for use in collecting anonymous population data to determine statistical relationships between well-being and other relevant variables, such as gender, health status and age.

The scale was developed from scrutiny of 435 adjectives and sentences derived from a review of the literature relating to wellbeing. An initial measure (Affectometer 1) comprising 96-items was tested in a general population sample and reduced to the 40 item Affectometer 2. Respondents are asked 'Thinking about the past few weeks, how often, if ever, have you felt...' and given six response categories (not at all, occasionally, some of the time, often, all of the time or don't know) followed by 20 positive items and 20 negative items each split into 10 sentences and 10 adjectives (table 1). The scale is scored by subtracting the sum of all negative items from the sum of all positive items, so that the total score has a potential range of -80 to +80.

Although the authors suggest that the tool is suitable for use in any English-speaking population, it has not been validated for use in the U.K.^{xii}. It has, however, been used to monitor the effectiveness of programmes or interventions at an individual level in other countries and had proved sensitive to change^{xiii,xiv}.

Methodology

Using data collected in the fourth wave of the Scottish Health Education and Population Survey (HEPS)^{xv} carried out in 2002, we assessed response rates to Affectometer 2 items, subscales and full scale and provide descriptive data, mean scale, subscale and item scores. These data are compared to data from the original New Zealand validation to assess whether response patterns were similar in both settings. The scale's internal consistency was tested using Cronbach's alpha and the underlying structure of the scale evaluated using principal components analysis. The correlation between positive and negative scale items was tested using Pearson's correlation coefficient. Prior hypotheses about the expected association between Affectometer 2 score and factors known to predict poor mental health were developed to test the tool's construct validity. Based on the findings of recent U.K. psychiatric morbidity studies^{xvi,xvii}, we hypothesised that men would show a higher score than women, that there would be no association with age at leaving full-time education and that the scale would show a positive association with higher socio-economic status.

Associations between self-reported health status and Affectometer 2 score were tested. These results were compared to a similar analysis of General Health Questionnaire (GHQ-12) score, carried out on the same dataset and the extent of correlation between GHQ-12 score and Affectometer score was assessed. We hypothesised that the scale would show a moderate correlation with the General Health Questionnaire. We also examined the performance of four different short versions of the scale: all 20 positive items; all 20 negative items; all 20 adjectives and all 20 statements.

NOT FOR WIDER CIRCULATION OR REPRODUCTION

Data collection

The HEPS survey gathered data from 1193 Scottish citizens aged 16 to 74 selected from the postcode address file for the Scottish mainland using random probability sampling. Subjects were contacted in person by a trained interviewer. A full interview was carried out at 72% of eligible addresses. The survey was administered by BMRB International for the Health Education Board for Scotland. The questionnaire included twenty sections covering subjects' self-reported health status, health behaviours and beliefs, socio-economic characteristics (sex, age, marital status, age left formal education, income of principal household earner and social grade) and two measures of mental health (GHQ-12 and Affectometer 2).

Statistical analyses

Data were analysed using SPSS for Windows (Version 13.0.1) software. Subjects that did not complete all 40 Affectometer items were excluded from our analyses. Cronbach's alpha scores were generated (negative items were reverse scored) and principal component analysis was carried out, using direct oblim rotation to be consistent with the scale developers' conception of the scale representing a single underlying concept of mental well-being. Construct validity was assessed using t-tests, one-way ANOVA and descriptive analyses.

Results

Sample characteristics

758 respondents out of 848 people who took part in the survey (91%) answered some or all of the Affectometer 2 questions with 89% of these (n=722) completing all forty Affectometer 2 questions (64% of those surveyed).

NOT FOR WIDER CIRCULATION OR REPRODUCTION

There were significant differences between responders and non-responders in mean age (responders 42 years, non-responders 48 years ($t = -2.86$, $p = 0.004$) but not gender. Non-respondents were more likely to be divorced or separated than respondents (14.8% compared with 10.2%) and were more likely to be in social grade D or E (33.1% compared with 29.5%). We observed similar differences between people who completed the GHQ-12 section of the survey and those who did not.

Non-completion rates for individual items ranged from 1.2% ($n = 9$) for the item 'I like myself' to 0 for three items ('Like a failure', 'Interested in other people' and 'I have energy to spare'). (table 1)

Non-completion rates for each of the four 20-item scales were compared. There was no significant difference between the four sub-scales ($p = 0.5$). (table 2)

Score characteristics

Scores ranged from -59 to 77, covering most of the available score range (-80 to 80). The mean score was 29.67 ($SD = 24.11$) with a median score of 33. Total scores for positive items ranged from 7 to 78 and negative items from 0 to 68.

Individual item means ranged from 1.45 to 2.92 for positive items and from 0.42 to 1.61 for negative items (table 3). Scores were similar to those observed in New Zealand population samples.

Principal components analysis

The Scree Plot method^{xviii} was used to determine the number of factors underpinning the scale. This method was used in preference to the Kaiser method (removing all components with eigenvalues under 1.0) because it draws on the relative rather than

NOT FOR WIDER CIRCULATION OR REPRODUCTION

absolute eigenvalues, and so is not so sensitive to the number of variables in the analysis^{xix, xx}, a potential problem with a scale the length of Affectometer 2. The Scree Plot method suggested a one factor solution (figure 1) in which positive and negative items are related, providing support for scoring the Affectometer 2 as a single score. In order to examine the common hypothesis that positive and negative factors are independent, we explored a forced two factor solution as well. In the two component model (table 4) items mostly showed moderate loadings for either positive or negative components. The one-factor solution explained 31% of total variance, with the two factor solution explaining 40%. To assess the extent of the possibility of two underlying scales we carried out an analyses of subscales comprised of positive and negative items separately.

Correlation between positive and negative items

Positive and negative items scores showed a moderate to high correlation ($r=0.604$).

Scale validity and reliability

The full scale had an internal consistency of 0.944 (negative items reverse-scored). All four twenty-item sub-scales (adjectives, sentences, positive items, negative items) performed comparably (table 5) and correlated well with the full scale. In particular the negative scale correlated -0.88 and the positive 0.92. Internal consistency was lower for the sub-scales than the full scale but was high, (greater than 0.9)^{xxi} for all sub-scales.

NOT FOR WIDER CIRCULATION OR REPRODUCTION

Construct validity

Statistically significant differences in mean scale scores ($p < 0.05$) were observed between all variables tested, with the exception of age at leaving full-time education. Similar associations were observed between these variables and GHQ-12 score, with the exception of age which did not show a statistically significant difference in GHQ-12 score between age bands (table 6).

Affectometer 2 correlated -0.600 with the GHQ-12 in line with anticipated performance.

Sub-scale performance

Analysis of the same variables by mean sentence score, mean adjective score, mean positive item and negative item score showed an identical pattern of associations for all variables.

Discussion

Although there was some responder bias in this sample, which showed lower response rates in older people, divorced people and people in lower socio-economic group, the extent of bias was very similar to that observed with the GHQ-12 suggesting that Affectometer 2 has a similar level of acceptability at a population-level to other comparable instruments. Response rates for both instruments may have been influenced by the inclusion of Affectometer 2 as part of a larger survey. If the tool is administered as a stand-alone measure, response rates may differ to those found in this study.

Individual item means were similar in this sample to those reported in the original validation survey in New Zealand (positive items means: 1.45 to 2.92 U.K. cf 1.64 to

NOT FOR WIDER CIRCULATION OR REPRODUCTION

2.96 New Zealand and negative item means: 0.42 to 1.61 U.K. cf 0.36 to 1.82 New Zealand), suggesting that the scale acts in very much the same way in the UK as it does in its country of origin. Since these data were collected we have carried out a qualitative study of UK respondents' understanding both of individual items and of the full score. Results, presented elsewhere^{xxii} confirm that members of the public in the UK thought the scale was measuring their mental and emotional wellbeing and that items were generally comprehensible and easy to respond to.

Unlike other commonly used measures of mental health, Affectometer 2 did not show a ceiling effect in this population, indicating that the measure has potential for documenting overall improvements in population health. Scores follow a distribution which although showing a negative skew is close to Normal and covers most of the available score range. Earlier studies showing that the measure is sensitive to change are also reassuring in this respect^{6,7}.

The results of our principal component analysis provide some support for the theory of a one factor solution. In this solution, all items loaded greater than 0.30 with 70% of items showing loadings greater than 0.5 and positive and negative items loaded in the expected way. This suggests that the items were related and can be used as recommended to derive a single score. However this single factor only explained 31% of the total variance. Those who believe in the independence of positive and negative mental health might be tempted to favour the forced two factor solution in which positive and negative items tended to load separately. This explained an additional 9% of the variance. Most items in the two factor solution showed loadings of greater than 0.5 on either the positive or negative. The correlation between positive and negative subscale was also moderate. Confirmatory factor analysis on

NOT FOR WIDER CIRCULATION OR REPRODUCTION

another data set would be helpful to confirm our finding that the 40-item scale reflects a single underlying construct.

Tests for construct validity showed significant differences in mean scores between men and women and between social classes (being male and higher socio-economic status was associated with higher Affectometer 2 scores) in line with expected hypotheses, derived from recent population surveys. Household income also showed a positive gradient: lower income levels were associated with lower Affectometer 2 scores. Mean scores decreased with worsening self-reported health status. No clear pattern emerged when the effects of age at leaving education were examined, which is also compatible with other findings⁶. Marital status was significantly associated with Affectometer 2 score with the highest mean scores for respondents who were married or in a relationship and lowest scores for those who were widowed or divorced as observed in some^{xxiii, xxiv} but not all mental health population surveys¹⁶. The level of correlation of Affectometer 2 with the GHQ-12, while showing the anticipated overlap, also confirms that the former is capturing something more than the absence of mental illness and low level of affect.

The high internal consistency scores for both the 40-item scale and four 20-item scales suggest that there is redundancy in both the full and reduced scales. The length of the full scale is a disadvantage and further development to shorten this measure would be of value. Assessment of the scales measurement properties using a RASCH analysis would also provide valuable information.

Conclusions

In this reasonably representative sample of 18-74 year olds in the UK population the Affectometer 2 performed well, showing similar response patterns to those observed

NOT FOR WIDER CIRCULATION OR REPRODUCTION

in the original New Zealand validation, high level of internal reliability and validity and expected correlations with socio-demographic factors. The scale discriminates between different populations in groups in a way that is consistent with the findings of other population mental health surveys.

The high level of internal consistency of both the full and reduced scales point to the opportunity of developing a shortened version of the scale which may be more acceptable for use as a population tool.

-
- ⁱ World Health Organisation Promoting Mental Health; Concepts emerging evidence and practice. Summary report. Geneva; World Health Organisation, 2004
- ⁱⁱ World Health Organisation. Strengthening mental health promotion. Geneva; World Health Organisation 2001.
- ⁱⁱⁱ Ryan RM, Deci EI. On happiness and human potential: a review of research on hedonic and eudaimonic well-being. *Ann Rev Psych*, 2001; 141-166.
- ^{iv} Goldberg DP, Williams P. A user's guide to the General Health Questionnaire. Windsor, UK: NFER-Nelson; 1988
- ^v Keyes CLM. Mental illness and / or mental health? Investigating axioms of the complete state model of health. *Journal of Consulting and Clinical Psychology* 2005; 73 (3): 539-548
- ^{vi} Huppert FA, Whittington JE. Evidence for the independence of positive and negative well-being: implications for quality of life. *British Journal of Health Psychology* 2003; 18 (1): 107-122.
- ^{vii} Department of Health. Our health, our care, our say: a new direction for community services. London: Department of Health, 2006.
- ^{viii} European Commission. Promoting the mental health of the population. Towards a strategy on mental health for the EU. COM (2005) 484. Brussels: European Commission, 2005.
- ^{ix} Department of Health. Choosing Health: making healthy choices easier. London: Department of Health, 2004.
- ^x Stewart-Brown S. Measuring the parts most measures do not reach: a necessity for evaluation in mental health promotion. *Journal of Mental Health Promotion* 2002; 1(2): 4-8.
- ^{xi} Kammann R, Flett R. A scale to measure current level of general happiness. *Aust Psychol* 1983; 35: 259-265.
- ^{xii} Kammann R, Flett R. Sourcebook for measuring well-being with Affectometer 2. 1983, Dunedin, New Zealand: Why Not? Foundation
- ^{xiii} Lo J, Zemke R. The relationship between affective experiences during daily occupations and subjective well-being measures: a pilot study. *Occupational Therapy in Mental Health*, 1997; 13(3); 1-21
- ^{xiv} Raeburn JM, Atkinson JM, Dubignon JM, Fitzpatrick J, McPherson M, Elkind GS. Superhealth basic: development and evaluation of a low-cost community-based lifestyle change programme. *Psychology and Health*. 1994; 9 (5): 383-95.
- ^{xv} Malam S, Angle H, Wimbush H, Fraser E. Health Education Population Survey 1996-2003., Edinburgh, NHS Health Scotland, 2004.
- ^{xvi} Singleton N, Bumpstead R, O'Brien M, Lee A, Meltzer H. Psychiatric morbidity among adults living in private households. Office for National Statistics: London, 2001.
- ^{xvii} Singleton HLG ed. Better or worse: a longitudinal study of the mental health of adults living in private households in Great Britain. The Stationery Office: London, 2003.
- ^{xviii} Cattell RB (1966). The scree test for the number of factors. *Multivariate Behavioural Research* 1966, 1, 140-161.
- ^{xix} Cooper, C. (1998). Individual differences. London: Arnold, 1998.
- ^{xx} Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432-442, 1986.
- ^{xxi} Bland JM, Altman DG. Statistics notes: Cronbach's Alpha. *BMJ* 199; 314: 572.
- ^{xxii} Stewart-Brown S, Tennant R, Fishwick R, Joseph S. Affectometer 2: a validation study. NHS Health Scotland, 2006.
- ^{xxiii} Regier DA, Farmer ME, Rae DS, Myers JK, Kramer M, Robins LN, George LK, Karno M, Locke BZ. One-month prevalence of mental disorders in the United States and sociodemographic characteristics: the Epidemiologic Catchment area study. *Acta Psychiatr Scand* 1993; 88 (1): 35-47.
- ^{xxiv} Smith CJ, Mercy JA, Conn JM. Marital status and the risk of suicide. *Am J Public Health* 1988; 78: 78-80.

NOT FOR WIDER CIRCULATION OR REPRODUCTION

Table 1: Analysis of missing data items

Item	% incomplete (n)	Item	% incomplete (n)
My life is on the right track	0.5 (4)	Satisfied	0.3 (2)
I wish I could change some part of my life	0.3 (2)	Optimistic	0.7 (5)
My future looks good	0.5 (4)	Useful	0.5 (4)
I feel as if the best years of my life are over	0.4 (3)	Confident	0.3 (2)
I like myself	1.2 (9)	Understood	0.5 (4)
I feel there must be something wrong with me	0.7 (5)	Interested in other people	0
I can handle any problems that come up	0.3 (2)	Relaxed about things	0.3 (2)
Like a failure	0	Enthusiastic	0.3 (2)
I feel loved and trusted	0.4 (3)	Good natured	0.1 (1)
I have been left alone when I don't want to be	0.4 (3)	Clear headed	0.5 (4)
I feel close to people around me	0.5 (4)	Discontented	0.5 (4)
I have lost interest in other people and don't care about them	0.4 (3)	Hopeless	0.3 (2)
I can do whatever I want to do	0.8 (6)	Insignificant	0.7 (5)
My life seems stuck in a rut	0.1 (1)	Helpless	0.4 (3)
I have energy to spare	0	Lonely	0.4 (3)
I can't be bothered doing anything	0.3 (2)	Withdrawn	0.3 (2)
I have been smiling and laughing a lot	0.4 (3)	Tense	0.3 (2)
Nothing seems very much fun anymore	0.3 (2)	Depressed	0.1 (1)
I have been thinking clearly and creatively	0.7 (5)	Impatient	0.1 (1)
My thoughts have been going round in useless circles	0.3 (2)	Confused	0.5 (4)

Table 2: Sub-scale response rate

Sub-scale	No. completing all 20 items	% complete response rate (of those who completed any Affectometer items)
Sentence scale	732	97%
Adjective scale	740	98%
Positive items	730	96%
Negative items	740	98%

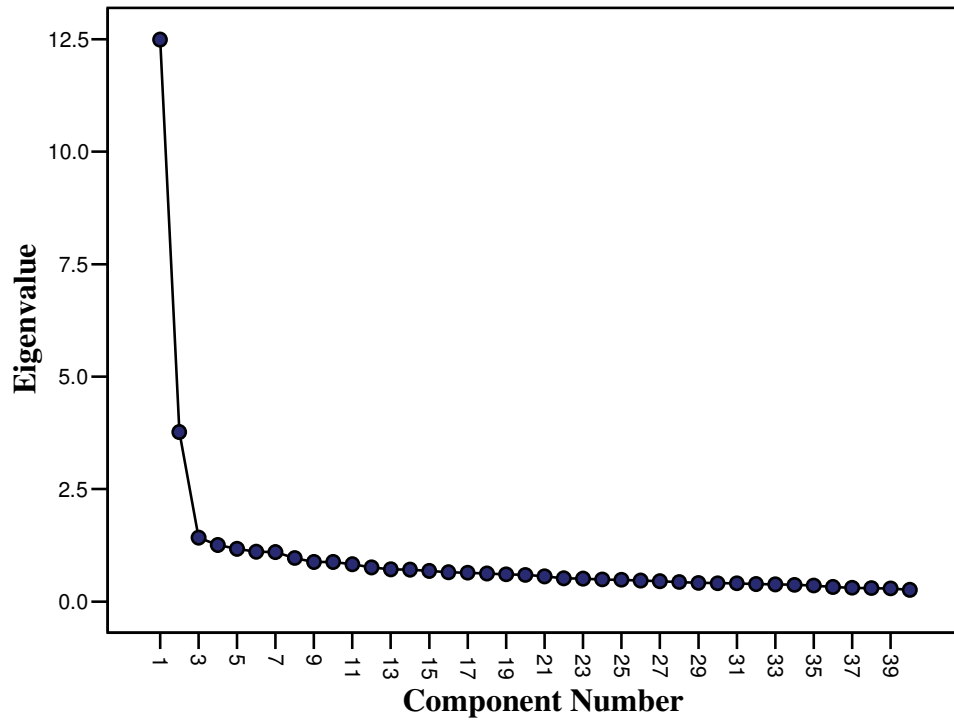
Table 3: Individual item means

Item	Mean (SD) (UK sample)	Mean (SD) (NZ sample)	Item	Mean (SD)	Mean (SD) (NZ sample)
My life is on the right track	2.31 (1.16)	2.60 (0.98)	Satisfied	2.21 (1.04)	2.52 (0.82)
I wish I could change some part of my life	1.57 (1.22)	1.82 (1.17)	Optimistic	2.11 (1.05)	2.54 (0.95)
My future looks good	2.16 (1.13)	2.42 (1.10)	Useful	2.42 (1.09)	2.58 (0.79)
I feel as if the best years of my life are over	0.78 (1.08)	0.66 (0.99)	Confident	2.41 (1.09)	2.32 (0.89)
I like myself	2.27 (1.13)	2.07 (1.07)	Understood	2.22 (1.03)	2.35 (0.95)
I feel there must be something wrong with me	0.64 (1.02)	0.76 (0.99)	Interested in other people	2.58 (1.03)	
I can handle any problems that come up	2.46 (1.09)	2.43 (1.14)	Relaxed about things	2.23 (1.00)	2.03 (1.12)
I feel like a failure	0.55 (0.95)	0.76 (0.85)	Enthusiastic	2.13 (1.02)	2.38 (0.98)
I feel loved and trusted	2.92 (1.08)	2.96 (1.14)	Good natured	2.60 (0.92)	2.73 (0.75)
I have been left alone when I don't want to be	0.49 (0.86)	0.84 (0.83)	Clear headed	2.44 (1.05)	2.65 (1.00)
I feel close to people around me	2.84 (1.01)	2.66 (1.03)	Discontented	0.95 (1.00)	1.13 (0.99)
I have lost interest in other people and don't care about them	0.42 (0.81)	0.36 (0.73)	Hopeless	0.45 (0.96)	0.54 (0.89)
I can do whatever I want to do	2.05 (1.21)	2.17 (1.03)	Insignificant	0.57 (0.89)	0.96 (0.99)

NOT FOR WIDER CIRCULATION OR REPRODUCTION

Item	Mean (SD) (UK sample)	Mean (SD) (NZ sample)	Item	Mean (SD)	Mean (SD) (NZ sample)
My life seems stuck in a rut	1.10 (1.16)	0.98 (1.04)	Helpless	0.55 (0.93)	0.73 (0.97)
I have energy to spare	1.45 (1.06)	1.64 (1.08)	Lonely	0.63 (0.98)	0.88 (0.95)
I can't be bothered doing anything	1.20 (1.00)	1.16 (1.00)	Withdrawn	0.62 (0.91)	0.96 (1.03)
I have been smiling and laughing a lot	2.30 (0.98)	2.71 (0.91)	Tense	1.35 (1.04)	1.34 (1.91)
Nothing seems very much fun anymore	0.76 (1.01)	0.88 (1.03)	Depressed	0.69 (1.01)	0.94 (0.94)
I have been thinking clearly and creatively	2.21 (1.02)	2.34 (0.89)	Impatient	1.62 (1.04)	1.71 (0.98)
My thoughts have been going round in useless circles	0.94 (1.12)	0.93 (0.76)	Confused	0.77 (0.98)	0.99 (1.08)

Fig. 1: Scree plot of Affectometer 2 eigenvalues



NOT FOR WIDER CIRCULATION OR REPRODUCTION

Table 4: Item loadings for one and two factor solution

	One factor solution	Two factor solution	
	Component 1	Component 1	Component 2
Your life is on the right track	-.60	-.39	.68
You want to change some part of your life	.56	.60	-.36
Your future looks good	-.57	-.35	.68
As though the best years of your life are over	.54	.59	-.32
That you like yourself	-.49	-.28	.60
There must be something wrong with you	.66	.73	-.36
You can handle any problems that come up	-.46	-.25	.59
Like a failure	.68	.76	-.38
Loved and trusted	-.49	-.30	.58
You have been left alone when you don't want to be	.39	.42	-.24
Close to people around you	-.56	-.36	.63
You have lost interest in other people and don't care about them	.39	.45	-.20
You can do whatever you want to do	-.42	-.21	.58
Your life seems stuck in a rut	.72	.76	-.47
You have energy to spare	-.28	-.15	.35
You can't be bothered doing anything	.52	.59	-.28
That you have been smiling and laughing a lot	-.54	-.34	.63
Nothing seems very much fun anymore	.65	.70	-.41
That you have been thinking clearly and creatively	-.59	-.37	.68
Your thoughts have been going round in useless circles	.67	.74	-.38
Satisfied	-.61	-.40	.69
Optimistic	-.48	-.27	.60
Useful	-.56	-.38	.61

NOT FOR WIDER CIRCULATION OR REPRODUCTION

	One factor solution	Two factor solution	
	Component 1	Component 1	Component 2
Confident	-.58	-.35	.70
Understood	-.51	-.29	.62
Interested in other people	-.37	-.21	.46
Relaxed about things	-.60	-.41	.66
Enthusiastic	-.52	-.31	.63
Good natured	-.49	-.31	.58
Clear headed	-.58	-.38	.65
Discontented	.63	.71	-.34
Hopeless	.66	.76	-.35
Insignificant	.61	.67	-.35
Helpless	.63	.71	-.34
Lonely	.59	.64	-.35
Withdrawn	.62	.71	-.32
Tense	.56	.62	-.32
Depressed	.73	.78	-.44
Impatient	.33	.40	-.16
Confused	.54	.62	-.28

Table 5: Sub-scale characteristics

Sub-scale score	Minimum score	Maximum score	Mean score	Standard deviation	Correlation with full scale	Internal consistency
Adjective	-31	40	15.1	12.1	0.96	0.91
Sentence	-32	38	14.5	12.6	0.97	0.90
Positive items	7	78	46	12.9	0.92	0.91
Negative items	0	68	17	13.4	-0.88	0.94

Table 6: Affectometer 2 scores (mean and standard error of the mean) by key variables

	Affectometer mean score (standard deviation)	GHQ-12 mean score (standard deviation)
Sex		
Male	37.8 (20.1)	1.3 (2.4)
Female	27.6 (27.3)	2.1 (3.4)
	p=0.019	p=<.001
Age bands		
16-24	23.6 (22.2)	1.8 (2.6)
25-34	27.3 (27.0)	2.0 (3.1)
35-44	32.9 (25.3)	1.7 (3.1)
45-54	27.0 (22.3)	2.0 (3.3)
55-64	33.9 (21.6)	1.4 (2.8)
65-74	36.7 (20.5)	1.1 (2.3)
	p=<.001	p=0.24
Marital status		
Single	25.4 (22.9)	1.7 (2.6)
Married/ living as a couple	32.5 (23.6)	1.5 (2.8)
Widowed/ divorced/ separated	23.7 (26.3)	2.9 (3.9)
	p=<.001	p=<.001
Self-perceived health status		
Very good	36.0 (20.4)	1.1 (2.2)
Good	30.8 (21.1)	1.6 (2.7)
Fair	15.9 (31.2)	3.0 (4.2)
Poor/ very poor	10.7 (29.2)	4.2 (4.1)
	p=<.001	p=<.001
Household income		
Under £5,000 pa	22.2 (24.8)	2.3 (3.2)
£5,000 to £14,999 pa	22.0 (29.8)	2.5 (3.8)
£15,000 to £29,000 pa	34.1 (23.5)	1.4 (2.7)
£30,000 or over pa	36.4 (17.0)	1.3 (2.3)
	p=<.001	p=<.001
Age at leaving full-time education		
16 or under	30.1 (23.7)	1.6 (2.9)
17-18	29.5 (23.2)	1.9 (3.2)
Over 18	28.9 (27.3)	2.1 (3.3)
	p=0.78	p=0.11
Social grade of chief income earner		
A (Upper Middle Class)	38.4 (18.9)	1.6 (2.5)
B (Middle Class)	33.0 (18.6)	1.4 (2.3)
C1 (Lower Middle Class)	32.1 (23.0)	1.2 (2.8)
C2 (Skilled Working Class)	31.0 (23.1)	1.3 (2.7)
D (Working Class)	26.8 (23.4)	1.5 (2.7)
E (Lowest level of subsistence)	13.8 (33.1)	3.4 (4.5)
	p=<.001	p=<.001

Appendix 3: Focus group information sheet

AFFECTOMETER 2 FOCUS GROUP INFORMATION SHEET - *version 3 (7/07/05)*

Study title: An assessment of an emotional well-being questionnaire - Affectometer 2

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Thank you for reading this.

What's this study all about?

Affectometer 2 is an emotional well-being questionnaire that has been used in a survey in Scotland. We want to talk to people about what they think of the questionnaire. We want to do this so that we can assess whether the questionnaire should be used in future UK surveys (England as well as Scotland). The study will finish in June 2006.

Why have I been chosen?

We are inviting different groups of people to take part in focus groups. This is so that we can compare whether people's age, being male or female or living in different places affects the way they answer the questionnaire. This will help us to assess whether the questionnaire can be used with everyone.

What would the research involve?

We are asking participants to take part in a focus group which will last about 1 hour. A focus group brings together 6-8 people to discuss an issue on which they are likely to have interesting views based on their own experience. The group will be led by a member of the research team. He will suggest the topics for discussion. There are no right or wrong answers. It's your opinions that count. You will also be asked to complete the questionnaire (but we won't ask you what answers you gave) and to comment on its content, format etc. You will be able to take your completed questionnaire away with you at the end of the focus group.

Would focus groups be recorded?

Yes. They will be tape recorded and transcribed. We need to do this so that we can make sure that our analysis is based on what people actually said in the focus groups. All recordings will be destroyed after use. Any quotations used will be anonymised so that no-one will be able to identify you.

NOT FOR WIDER CIRCULATION OR REPRODUCTION

Will I get paid for taking part?

No. We will however pay each participant some money to cover any expenses such as bus or train fares that you have had to pay to come to the focus group. If you need any help with child care or other caring responsibilities, we can also help with this: please talk to the person who has given you this information sheet for more information.

Will what I say affect the care I get from the NHS or social services?

No. The study is being paid for by NHS Health Scotland but there is no connection between this study and any care you might be receiving now, or will receive in the future, from the NHS or social services.

Who is paying for this research?

This research is being paid for by NHS Health Scotland.

Could taking part in this study do me any harm?

Talking about emotional well-being can be a positive experience but it may also raise difficult issues that you may feel upset or uncomfortable about. If you decided to take part and you feel upset or uncomfortable or just want some time out from the discussion, you may leave the room at any point. The project researcher, Ruth Fishwick will be on hand to help you during the focus group, if you want. If you would like to talk to someone after the focus group, here are some contact numbers for organisations that you can talk to in confidence:

SANELINE - 0845 767 8000 (calls charged at local rates)

Samaritans - 08457 90 90 90 (calls charged at local rates) or you can email jo@samaritans.org

Can I change my mind if I volunteer now but have second thoughts?

Yes. You can change your mind about taking part in the study at any time, including during the focus group.

What will happen to the results of the research study?

The results of this research will be given to NHS Health Scotland. They will use this information to decide whether to use the questionnaire in national surveys in the future. The results will also be published in journals and presented at conferences.

Where can I get more information from?

If you would like to know more about this study or have any other questions, you can telephone Ruth Fishwick on 02476 575 772 or you can email ruth.fishwick@warwick.ac.uk

If you decide to take part in this research you will be given a copy of this information sheet to keep and will be asked to sign 2 consent forms, 1 for you and 1 for Ruth Fishwick.

Thank you for your time!

Appendix 4: Focus group consent form

Study centre: Warwick Medical School

Study ID Number: 2004/2005 RE053

Participant Identification Number for this trial:

CONSENT FORM- version 2 (7/07/05)

Title of Project: An assessment of an emotional well-being questionnaire - Affectometer 2

Name of Researcher: Ruth Fishwick

1. I confirm that I have read and understand the information sheet dated 7/07/05 (version 3) 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 medical care or legal rights being affected.
3. I understand that the interview will be tape recorded
4. I understand that anonymous quotes from interviews may be used by the researchers in publications. These quotes will always be anonymous and it will not be possible to identify people who took part in the focus groups from the quotations
5. I agree to take part in this study.

Please put your initials in these boxes.

Name of Participant

Date Signature

Name of Person taking consent
(if different from researcher)

Date

Signature

Researcher

Signature

Date

Please keep one copy of this form and give the second copy to Ruth Fishwick.

Appendix 5: Affectometer 2 - Schedule for Focus Groups

Version 3 (amended following pilot focus groups) (7/7/05)

Introductions & consent process

Give out name badges & introduce everyone.

Ruth Fishwick (RF) to go through information sheet and consent form again (participants will have received these at least a week before the focus group and been briefed by the community group leader or contact), giving participants the opportunity to re-read the information sheet, ask questions etc. Reinforce that participants are free to leave at any point.

Participants to sign 2 consent forms, keep 1 and return the other to RF. Participants to keep the information sheet.

Main session

Explain how session is going to run. Emphasise – not a test – no right or wrong answers – it's their views which we want to hear and are important

Affect. handed out – respondents asked to complete – no detail given at this point about what it is attempting to measure – we want a blank canvas in order that respondents can give an unbiased view

Respondents assured that that can keep the completed Affect. – they will not have to hand it in to the researchers or explain any of their answers

a) What were your first thoughts/feelings when you saw this (emphasise again your FIRST thoughts/feelings)?

Prompt on any initial thoughts respondents had on purpose of Affect.
Capture here and throughout – language issues – do people display positive/negative reactions to any of the words/language used?

b) Ok – moving on from your first thoughts/feelings – thinking about when you were reading through the questions and filling it in – what did you think/feel about it?

c) How did you find it to complete – and I'm thinking here about how easy/difficult or not you found it to fill in?

Expand with respondents on why/why not it was easy/difficult to complete
Do people find it easier to complete the sentence chunk of the questionnaire or the single word chunk?

d) What do you think the questions/questionnaire was trying to get at?

Only after they have offered their own ideas:

Prompt on:

Mental health

Positive mental health

NOT FOR WIDER CIRCULATION OR REPRODUCTION

Well-being

If none of the prompted terms appear appropriate - tease out what these terms suggest to them e.g. is "mental health" taken to be "mental ILL health"?

Prompt on if these terms/phrases ring true for people – i.e. do they make sense or do they not like them – what do people associate with these phrases?

Tease out anything which participants mention which they think are important parts of their mental well-being that aren't included in the Affectometer

PARTICIPANTS SCORE THEIR OWN QUESTIONNAIRE AT THIS POINT

Scores to be compared to distribution of scores from HEBS population survey (poster on wall). Explain that in a survey done across the whole of Scotland, most people scored between 0 and 30. Explain that your score can change a lot depending on what's been going on in your life recently.

e) Now that you've scored it – what do you think of your own score?

Does the score seem to be "about right" for yourself – or is too high/low?

Tease out anything which participants mention which they think are important parts of their mental health that aren't included in the Affectometer

If it's possible to agree on a term which signifies good mental health and well being – do they think this state is the opposite of mental illness/ill health (i.e. well-being implies an absence of mental illness) OR can actually co-exist (i.e. you can be mentally ill yet enjoy mental well-being)

Throughout the discussion issues that respondents raise re. language used in the Affect. would be noted and briefly expanded upon – but would not become the focus of the discussion given the length of the Affect. and the fact that other issues need to be covered in a limited amount of time.

Co-worker/RF to note any embarrassment/hostility etc that may arise relating to the Affect. – especially those things which the tape "can't see"

End of session – ask how people found the session. Emphasise that if anyone would like to speak to the moderator or researcher confidentially after the session they can do so. Moderator & co-worker to hand out leaflets with helpline numbers on to participants.

Close with thanks

Answer any questions

Explain that will give feedback through the group organiser

Expenses to be handed out & signed for by participants

Appendix 6: Affectometer 2 questionnaire information sheet

Version 2 (7/07/05)

Study title: An assessment of an emotional well-being questionnaire

Affectometer 2

You are being invited to take part in a research study being run by Warwick & Edinburgh Universities. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Take time to decide whether or not you wish to take part.

Thank you for reading this.

What is the purpose of the study?

This study is testing whether Affectometer 2 is a good questionnaire for measuring people's emotional well-being and how well it compares to other similar questionnaires. We want to do this so that we can assess whether Affectometer 2 could be used in national surveys in Scotland and England. The study will finish in June 2006.

Would I get paid to take part?

No: there is no payment for taking part in this study.

Why have I been chosen?

We are testing the questionnaire (and questionnaires that have a similar purpose) with four groups of people registered for under-graduate or post-graduate degrees at Warwick and Edinburgh Universities. Different groups of students have been selected to cover a range of academic disciplines and ages.

What does the research involve?

Ruth Fishwick will attend your teaching session next week. If you decide to take part, the researcher will ask you to sign 2 consent forms after your teaching session. Once you have completed these, Ruth Fishwick will give you 2 questionnaires to complete. If you can, please complete these straight away and return them to the researcher. If you cannot complete them straight away, please complete them as soon as you are able and return them to the researcher using the return envelope provided.

One week later, a randomly selected group of students from your group will be asked to fill in Affectometer 2 again.

NOT FOR WIDER CIRCULATION OR REPRODUCTION

Do I have to take part?

No. Participation in this research is completely voluntary.

What happens if I don't want to take part?

If you do not want to take part, you may leave your teaching session before the researcher hands out the consent forms.

Who is paying for this research?

This research is being paid for by NHS Health Scotland.

What will happen to the results of the research study?

The findings of this research will be given to NHS Health Scotland. They will use this information to decide whether to use Affectometer 2 in national surveys in the future. The results will also be published in journals and presented at conferences.

We will post a summary of the results from your group and a comparison of your group with the other teaching groups taking part in this survey on the notice board in your department. All data will be fully anonymised: no individual person or their responses to the questionnaire will be shown, only the average scores.

Where can I get more information from?

If you would like to know more about this study or have any other questions, please speak to the researcher attending this teaching session.

What do I do next?

If you decide to take part in this research, you will need to complete 2 consent forms (one for you and one for the project researcher) at next week's teaching session. You will then be given the 2 questionnaires to complete.

Where do I get more information about this research from?

If you have any other questions about this research, please contact Ruth Fishwick by email (ruth.fishwick@warwick.ac.uk) or be telephoning 02476 575 772

Thank you for reading this information sheet.

Appendix 7: Student Consent form

Study centre; Warwick/ Edinburgh University
Study ID Number: 2004/2005 RE053
Participant Identification Number for this trial:



CONSENT FORM *version 2 (7/07/05)*

Title of Project: An assessment of an emotional well-being questionnaire -
Affectometer 2

Name of Researcher: Ruth Fishwick

1. I confirm that I have read and understand the information sheet dated 7/07/05 (version 2) 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 medical care or legal rights being affected.
3. I agree to take part in this study.

Please put your initials in these boxes.

Name of Participant

Date Signature

Name of Person taking consent
(if different from researcher)

Date

Signature

Researcher

Signature

Date

Please keep one copy of this form and give the second copy to Ruth Fishwick.

Appendix 8: The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS)

Below are some statements about feelings and thoughts.

Please tick the box that best describes your experience of each over the last 2 weeks

STATEMENTS	None of the time	Rarely	Some of the time	Often	All of the time
I've been feeling optimistic about the future	1	2	3	4	5
I've been feeling useful	1	2	3	4	5
I've been feeling relaxed	1	2	3	4	5
I've been feeling interested in other people	1	2	3	4	5
I've had energy to spare	1	2	3	4	5
I've been dealing with problems well	1	2	3	4	5
I've been thinking clearly	1	2	3	4	5
I've been feeling good about myself	1	2	3	4	5
I've been feeling close to other people	1	2	3	4	5
I've been feeling confident	1	2	3	4	5
I've been able to make up my own mind about things	1	2	3	4	5
I've been feeling loved	1	2	3	4	5
I've been interested in new things	1	2	3	4	5
I've been feeling cheerful	1	2	3	4	5

“Warwick Edinburgh Mental Well-Being Scale (WEMWBS)
© NHS Health Scotland, University of Warwick and University of Edinburgh, 2006, all rights reserved.”