Overview of ONS phase three cognitive testing of Subjective Well-being Questions

In addition to the quantitative testing of subjective well-being questions on the Opinions and Lifestyle survey a qualitative research programme was integrated into ONS’ research strategy. This has had three phases to date:

2. Cognitive testing of the four subjective well-being questions, preambles, and alternative subjective well-being questions.
3. Further cognitive testing of the four subjective well-being and preamble as well as alternative wording and preamble for the four questions building on previous findings.

This summary report documents the third phase of testing which took place during October 2012.

Sample

The sample included 30 respondents with wide range of ages and educations levels. It also included respondents with a range of employment statuses (employed, economically inactive, and unemployed seeking work), and contained even numbers of men and women (see Annexe 5).

The priorities for cognitive testing in phase three testing were:

- The time periods respondents were using when answering questions.
- How respondents calculated their responses in relation to the 11-point scale.
- Interpretation of alternative terms, including worthwhile versus purpose and meaningful, and anxious versus worried and stressed.
- The effect on question three (happy yesterday) and question four (anxious/worried/stressed yesterday) of specifying that respondents should
think of yesterday, even if it was not a typical day, and the placement within the questions of that statement.

- How respondents felt about answering the questions in front of other household members and possible alternatives to reduce any possible response bias that may be associated with this.

The four overall monitoring questions that were included on the APS were:

- Overall, how satisfied are you with your life nowadays?
- Overall, to what extent do you feel the things you do in your life are worthwhile?
- Overall, how happy did you feel yesterday?
- Overall, how anxious did you feel yesterday?

All answered using a 0 to 10 scale where 0 is 'not at all' and 10 is 'completely'. In this report the abbreviations 'life satisfaction', 'worthwhile', 'happy yesterday' and 'anxious yesterday' are used to refer to the four overall monitoring questions outlined above.

**Time-periods for the ‘life satisfaction’ and ‘worthwhile’ questions**

**Background**

The ‘life satisfaction’ and ‘worthwhile’ questions do not have explicit time-frames. Instead, for the life satisfaction question the term ‘nowadays’ was included to focus the respondents on more recent times in their lives. This term was explored in relation to the question as well as an alternative term ‘these days’, a version of the questions with no time-frame, and a version which specified respondents to think about ‘the past year’. See Annexe 1 for a list of the four versions of life satisfaction questions tested. Variations on the worthwhile question which asked respondents to think about ‘the past year’ were also explored.

- People did not see any clear and consistent differences between ‘these days’ and ‘nowadays’ and when asked to explain if there was a difference, struggled to do so.
The inclusion of ‘nowadays’ sometimes had the effect of making the response more focused on the present, particularly when compared to a version of the question with no ending.

Respondents related ‘nowadays’ to particular events, either economic or political, or personal events significant in their own lives.

When the past year time frame was included in the life satisfaction and worthwhile questions, responses had a more consistent focus on life evaluation over a more similar period.

How respondents arrived at their rating on the 11 point scale and what they considered to be ‘high’, ‘low’, and ‘average’ ratings

- When asked what an ‘average’ rating was findings were mixed. Some respondents considered five to be neutral/average, others provided a range of scores that they considered to be within the average range, such as four to six, five to seven, seven to eight and three to seven.

- When asked what they would consider to be a ‘high’ score, views were mixed, although all were over five, many from seven upwards.

- What respondents considered to be a ‘low’ rating ranged from six and under, predominantly under five, some respondents giving scores as low as ‘under two’.

- When arriving at their own rating on the 0 – 10 scale. Some respondents took into account a number of different factors or life experiences and balanced these when answering arriving directly at a rating on the 0 – 10 scale, whereas others first worked out what they considered to be a neutral score before assessing how their personal rating should differ from the neutral rating then arrived on a point on the 0 – 10 scale.

- It should be noted that the survey questions take an average of 90 seconds to complete. It is therefore unclear how much people are able to make very detailed calculations in their scoring in an interview setting as compared to the cognitive interview which lasts substantially longer.
The ‘Worthwhile’ question

Background

Phase two testing found that some respondents with lower levels of education struggled to understand the term ‘worthwhile’. In phase three testing two alternative terms were tested, ‘purpose’ and ‘worthwhile’ (see annexe two).

- Purpose was understood by respondents in phase three, however, for some the term had negative connotations, whereas for others it had religious connotations.
- The term meaningful was understood, although younger respondents were unsure whether personal meaning or societal meaning should be considered.

‘Happy yesterday’ and ‘anxious yesterday’ questions

Background

In phase two of cognitive testing it was found that a group of respondents did not want to give a rating based on their emotions the previous day as it was not a reflection of their general state. Instead these respondents would provide a rating based on their average levels of happiness/anxiety as they felt that this better reflected their emotions more generally. In order to resolve this an instruction was introduced in phase two testing to ‘please think about yesterday, even if it was not a typical day’ and tested in a final wave of interviews. The instruction removed this issue, and respondents provided an answer based on their feelings the previous day rather than giving a rating which reflected their general state.

In phase three cognitive testing two different versions of the happiness question were tested; one version included an instruction to think about yesterday even if it was not a typical day, within the question. A second version was also tested which included the instruction as a preamble to both the happiness and anxious questions (see annexe 3).

- The yesterday instruction proved to be effective in ensuring that respondents considered only the previous day when arriving at their answer. Probing revealed that this was due respondents feeling reassured that answers could
be given even if the previous day was not typical, for example, providing a low score for the ‘happiness yesterday’ question would not imply that respondents were generally unhappy in their lives, this helped respondents provide an honest rating of their emotions the previous day.

- No clear difference in response was observed for the two versions of the instruction.

**Anxious yesterday question**

**Background**

Phase two cognitive testing found that respondents had a good understanding of the term ‘anxious’, however, there was some evidence that there was some level of stigma attached to the term ‘anxiety’ which could lead to deflated scores for this question. In addition, for the ‘anxiety’ question the scale is reversed, for example, a score of nought out of ten would indicate ‘no anxiety’, and is therefore positive. For the other three subjective well-being questions a score of nought out of ten would be the most negative result indicating, for example, that the respondent had ‘no satisfaction’ in their lives. This led to some respondents confusing the scale for the ‘anxious yesterday’ question and providing a high rating out of 10 when they had intended to provide a rating indicating low anxiety.

- Anxious’, ‘stressed’ and ‘worried’ are interchangeable terms for some respondents, but have very different meanings for others. Some respondents associated ‘anxiety’ with mental illness (see annexe 4).

- Both the terms ‘anxious’ and ‘stressed’ had both positive and negative connotations for these respondents, whereas ‘worried’ was only found to have negative connotations.

- There was evidence that the scale continues to be reversed by some respondents.

**Answering in front of others**

The APS is a household survey consisting of a combination of telephone and face-to-face interviews. Where face-to-face interviews are carried out it is possible that
other members of the household would be present at the time of the interview. These issues were explored with respondents during this phase of cognitive testing.

- Respondents were mixed in terms of whether they would mind answering the subjective well-being questions in front of other members of the household. Some felt that the presence of others would not affect their answer whereas others would be uncomfortable giving ‘low’ scores in front of partners, children, or parents, therefore answering in front of others may lead to inflated scores.

- Some female respondents stated that they would prefer to be interviewed by a woman as they would feel more comfortable thinking through personal issues and providing a rating with a women interviewer.

- There was also evidence that the demeanour of the interviewer asking the well-being questions would affect the respondent. Respondents stated that they would feel more “comfortable” providing a truthful answer if a “friendly” interviewer built rapport with them before asking the questions.

**Chart 1** provides an overview of the key findings.
**Chart 1: Overview of Key Findings, Stage 3 Cognitive Testing of SWB**

### Question time reference periods

- Open time reference periods such as ‘nowadays’ and ‘these days’ in the life satisfaction question are defined in very different ways by respondents and may lead to unpredictable response variation.
- Setting a more defined time period such as over the past 12 months (giving a month and year starting point) could lead to greater consistency and comparability.
- Providing more detailed instructions about time periods such as ‘Please think about yesterday even if it was not a typical day’- seems to encourage more consistent responses and reduces the temptation for respondents to re-interpret the question.

### Variations on worthwhile: purpose and meaning

- In Phase 2 testing, respondents with lower educational attainment had difficulty understanding the term ‘worthwhile’. This finding was not supported in Phase 3 testing.
- The term ‘purpose’ had religious connotations for some respondents but not for others suggesting possible inconsistency in interpretations of the question. The response scale for the ‘purpose’ question includes the word ‘purposeful’ which was not always viewed as equivalent to having purpose and the term was not understood by all respondents.
- The term ‘meaningful’ was well-understood, but led to questions of interpretation about ‘meaningful to whom’? For some, this was interpreted as meaningful to society whereas others interpreted it as personally meaningful. This was not noted with the ‘worthwhile’ question in this sample, but could be an issue which emerged there too.

### Variations on anxious: worried and stressed

- Anxious was viewed as more serious than either ‘worried’ or ‘stressed’ and was associated by some with mental illness. This could lead to under-reporting.
- In this sample, ‘worried’ was viewed as a negative emotion whereas both ‘anxious’ and ‘stressed’ were also sometimes also viewed as positive emotions (eg, nervous excitement).
- The response scale for the ‘anxious’ question continues to be reversed by some respondents and this is unlikely to be affected by these wording changes.

### Respondent methods for working out responses

- A variety of response methods were described including working down from the top of the scale (ie, deducting points for negative life circumstances), or working either up or down from a perceived scale midpoint (or mid range) taking into consideration both positive and negative circumstances.
- It is unclear how much these methods are used in practice in a series of questions lasting on average 90 seconds, but they suggest respondents tend to either compare their current life to their ‘ideal life’ or alternatively to work out how their life circumstances relate to a perceived middle ground.
- The score range for the ‘middle ground’ varied considerably, but ‘high’ scores were always considered to be over 5 among this sample.

### Answering in front of others

- The issue of possible response bias was explored in relation to respondents answering these potentially sensitive questions in the presence of other household members.
- Responses were mixed but indicated that some respondents would feel uncomfortable giving more negative scores in front of partners, parents or children and this could lead to inflated scores.
- Interviewer characteristics were also noted as possibly influencing responses.
Annexe 1

The current APS subjective well-being question on life satisfaction:

Overall, how satisfied are you with your life nowadays?

The alternative versions tested in phase 3 were:

Version A

Overall, how satisfied are you with your life these days?

Version B

Overall, how satisfied are you with your life?

Version C

During the past year, overall, how satisfied have you been with your life?

Annexe 2

The current APS subjective well-being question on worthwhile:

Overall, to what extent do you feel that the things you do in your life are worthwhile?

The alternative versions tested in phase 3 were:

Version A

Overall, to what extent do you feel that the things you do in your life have purpose?

Version B

Overall, to what extent do you feel that the things you do in your life are meaningful?
Annexe 3

Version A instruction – within the question

Overall, how happy did you feel yesterday? Please think about yesterday even if it was not a typical day.

On a scale where nought is ‘not at all worried’ and 10 is ‘completely worried’, overall, how worried did you feel yesterday? Please think about ‘yesterday’ even if it was not a typical day.

Version B instruction – before both affect questions

The next two questions ask about how you felt yesterday. Please think about yesterday even if it was not a typical day. Overall how happy did you feel yesterday?

On a scale where nought is ‘not at all stressed’ and 10 is ‘completely stressed’, overall, how stressed did you feel yesterday?

Annexe 4

The current APS subjective well-being question on anxious yesterday:

Overall, on a scale of nought to ten, how anxious did you feel yesterday?

The alternatives tested in phase 3 were:

Version A

On a scale where nought is ‘not at all worried’ and 10 is ‘completely worried’, overall, how worried did you feel yesterday? Please think about ‘yesterday’ even if it was not a typical day.

Version B

On a scale where nought is ‘not at all stressed’ and 10 is ‘completely stressed’, overall, how stressed did you feel yesterday?
**Annexe 5: SAMPLE COMPOSITION**

<table>
<thead>
<tr>
<th></th>
<th>Version A</th>
<th>Version B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age group</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 - 19 years</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>20 - 64 years</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>65+ years</td>
<td>6</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>16</td>
<td>14</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Educational Attainment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCSE or equivalent &amp; lower</td>
<td>3</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>‘A’ level or equivalent</td>
<td>5</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Degree or equivalent &amp; higher</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>N/A (aged 16-19)</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>16</td>
<td>14</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>8</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>Female</td>
<td>8</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>16</td>
<td>14</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Employment Status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>5</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Mode of administration of well-being questions</td>
<td>Telephone</td>
<td>Face-to-face</td>
<td>Total</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>-----------</td>
<td>-------------</td>
<td>-------</td>
</tr>
<tr>
<td>Economically inactive</td>
<td>8</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>Unemployed seeking work</td>
<td>3</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>14</td>
<td>30</td>
</tr>
</tbody>
</table>