National Research and Development Centre
for adult literacy and numeracy

PROPOSED SKILLS FOR LIFE SURVEY 2008

OPTIONS PAPER: NO. 2

Authors:

John Bynner, Greg Brooks, Diana Coben, Harvey Mellar, Augustin de Coulon, Samantha Parsons, John Vorhaus

2008

1. INTRODUCTION

This paper is intended to accompany its earlier and more detailed companion: Proposed Skills for Life Survey 2008 – Options Paper: Draft. With the exception of section 3, in which we specify the development work we recommend and the rationale for that, much of the methodological detail previously presented is omitted here.

In section 2 we provide the context for our recommendations, setting out the requirements, constraints and resources we took account of. Section 3 sets out our primary proposal; that is, a pre-pilot evaluation of the instruments to be taken forwards into the piloting phase of the 2008 Skills for Life Survey. Sections 4-6 provide further details of the recommended development work in relation to literacy, numeracy and ICT. In section 7 we briefly discuss the piloting phase (which we presume will be undertaken on a CAPI instrument developed by the field work agency involved in the survey) before summarising our primary recommendations in section 8.

The material is organized as follows:

Section 2: The context - requirements, constraints and resources

Section 3: Recommended development work

Section 4: Literacy - additional recommendations

Section 5: Numeracy - additional recommendations.

Section 6: ICT – recommendations

Section 7: Piloting and using the 2004 literacy and numeracy assessments

Section 8: Summary of key recommendations.

We provide our recommendations together with estimated costs, and comment on areas identified by DIUS as of particular importance:

- Spiky profiles
- Writing skills
- Numerical problem-solving
- ICT skills

2. THE CONTEXT: REQUIREMENTS, CONSTRAINTS, RESOURCES

Requirements of the new test

The critical requirement is to identify items to form an assessment following the same format as the BCS70 survey that:

1. enable the SfL original assessment to be replicated in terms of item properties (facility, discrimination, dimensionality) and the validity of the assessment as a whole
2. enable the BCS70 scale to be replicated with the same properties
3. can be used in the ongoing longitudinal Army study assessment.

The new test should enable us to:

1. determine whether performance in literacy and numeracy has improved or deteriorated at the different NQF performance levels (Entry 1, Entry 2, Entry 3, Level 1, Level 2) since the last survey carried out in 2003
2. identify 'spiky profiles'

In addition space will be released in the interview session to:

3. assess ICT competence
4. assess writing as well as reading, as measures of literacy performance.

Constraints

To avoid over-burdening respondents and to ensure that the additional assessments and relevant classificatory variables can be included, time must be saved on the literacy and numeracy assessment to release interview space.

Resources

1. The original Skills for Life (SfL) test as used in 2003, comprise a pool of 48 numeracy items with 19 attempted, and 40 literacy items with 25 attempted. These are used in the 'adaptive testing' format, whereby respondents complete items they are able to do up to a certain level and when they fail to achieve a 'pass' criterion switch to another dimension of the test.

2. The reduced form of the SfL test, used in the BCS70 age 34 survey will comprise 30 literacy and 17 numeracy items selected to tap the different NQF performance levels. After unsatisfactory results from pre-testing piloting the adaptive testing format was dropped for numeracy and respondents were given the opportunity to attempt all the items. . For literacy, adaptive testing was restricted to including a 'screen' to be sure that respondents had sufficient literacy capability to attempt the highest two levels of the test. This screen comprised ten items pitched at Entry 3 level. Respondents who failed five or more of these items dropped to items designed at Entry Level 2 and 1. Respondents who passed six or more of the items progressed to 5 Level 1 and 5 Level 2 items. A maximum of 20 items were attempted by all respondents. The national distributions for both the numeracy test and the literacy test were comparable to those of the SfL survey.

3. It has been agreed that the BCS70 literacy and numeracy tests will be piloted for use in the army survey.

3. RECOMMENDED DEVELOPMENT WORK: OVERVIEW

In this section we describe the aims and objectives of the development work we propose, including pre-piloting of the test. We distinguish between this and the piloting phase, which we briefly discuss in section 7, and which in any case will be undertaken on a CAPI instrument developed by the field work agency involved in the survey.

In sections 4 – 6 we provide additional detail of what the development work will comprise in relation to literacy, numeracy and ICT.

Test evaluation

We recommend that through analysis of the raw test data the measurement properties of the literacy and numeracy items across three different target populations or survey research settings should be determined and the tests evaluated for the purposes of the new survey. Namely:

- The original SfL 2003 assessment
- The BCS70 scale for respondents aged 34 years (and born in 1970)
- The pending Armed Services Study
- An alternative numeracy test based on a new selection of SfL items (p. 6)

Aims

The aim of this evaluation is to confirm the measurement properties of the original items used in the 2003 SfL survey in order to demonstrate the robustness of the literacy and numeracy scales, as replicating the key properties of the original SfL test.

Ideally, scales for literacy and numeracy should enable researchers to ‘add up’ the number of correct responses in order to determine a level of attainment. If a particular group of items can be distinguished as belonging to particular sets or levels of attainment then it is possible to assess whether or not these levels match on to national standards, and enable the researcher to present the items to the respondent as a series of ‘stepping stones’. Thus, if the first stone (set) is ‘wobbly’ underfoot then respondents remains at that position. Alternatively, if their footing is secure they advance to the next stone (set), and so on. In this way tests are described as ‘adaptive’, consisting of sets of items which are naturally easier or more difficult than others. Respondents proceed until they reach their natural level or ceiling. Provided the number of correct responses can be added up in a meaningful way individuals can be placed in order or ranked according to their abilities.

Essentially, the process of addition implies that each item can be thought of as informing the user about achievement in a particular domain of literacy or numeracy. These domains are referred to as dimensions. Thus, items belonging to

a single dimension are said to be unidimensional and all measure the same underlying phenomenon.

Confirmatory factor analysis enables the researcher to explicitly test this assumption and competing versions of the underlying dimensionality can be tested accordingly and the confirmed test structure replicated in other datasets. Typically, the initial choice of the number of dimensions is driven by a theory of measurement. In our case that there is a single concept, literacy or numeracy, that explains the observed correlations between the items. Empirical evidence is obtained to allow the researcher to make a decision as to whether or not to retain any of the items as belonging to a scale or level. Additionally, as a result of applying factor analysis some items may contribute more information or 'weight' to a scale score than others. In this way the analysis allows the researcher to score a performance along an underlying trait in a way that assigns greater or lesser numerical importance to individual items. The resulting factor score may not always provide the same rank ordering of performance as use of a raw summated score. However, it must be said that the routine computation or use of factor scores does not have widespread intuitive appeal when compared to creating a simple total of correct responses.

Secondly, an evaluation can provide the means by which an individual item can be said to discriminate between individual performances. In this way the observed probability of getting an item 'correct' should relate to an individual's total score. Individuals with higher scores are expected to have a higher probability of endorsing an item. In this way items which are only endorsed by individuals with high scores are said to discriminate between high and low achieving individuals. Items which are likely to be always correct across a range of varying total scores are not particularly efficient at discriminating between individuals and would as a result of an evaluation be dropped from a scale.

Objectives

We would build on this work in order to achieve the following objectives:

1. Strengthen the psychometric properties of the scales across three settings
2. Assess the difficulty/facility of each item
3. Assess the extent to which the relevant dimension can be measured reliably while reducing the number of items in a particular scale
4. Test the extent to which the properties of the scales are invariant for key subgroups in the population, e.g. for gender and ethnicity.

The resulting outcomes would be to produce a reliable set of literacy and numeracy scales that could be applied in a variety of national contexts. The timescale would consist of a 2 month programme of work for a research officer and a consultant statistician.

Schedule

Week	Research Officer	Consultant Statistician
1	Familiarisation: datasets, BCS70 & SfL 2003 Test software (MPlus, SPSS)	2 days advice, set-up, on the job training
2	BCS70 re-analysis using item response theory	2 days advice & direction
3	BCS70 subgroup analysis	1 day advice and direction
4	SfL2003 analysis	1 day advice and direction
5	SfL2003 subgroup analysis	1 day advice and direction
6	Reflect, write up, recommend	2 days help and advice with write up and recommendation
7	Armed Services pilot test ; confirm scale properties	2 days advice and direction
8	Write up report on measurement properties	2 days advice and direction
Totals	**8 weeks**	**13 days spread across 8 weeks** **Estimated total cost: £9,000**

4. LITERACY: ADDITIONAL RECOMMENDATIONS

Spiky profiles

We previously recommended that no attempt should be made to devise separate tests or items for spiky profiles in literacy ore numeracy. If spiky profiles exist they will emerge in the data. **Extra cost: nil.**

Writing

We recommend that the writing assessment within the set of Go! tools developed by NFER for NRDC be used. This has 2 parallel forms (only one would be needed for this survey), with 3 tasks within each. Sue Grief and Bill Meyer should be consulted on how they used the full instruments in the 'Effective Practice in Writing study' (in particular, how they modified the marking instructions provided by NFER), and on whether it is necessary to use all 3 tasks. In the 'Improving the Quality of Teaching and Learning in adult literacy project' Maxine Burton and colleagues at Sheffield University used only the first task in each form, namely (we paraphrase): 'Look at Go! for 10 minutes, then write a few sentences telling us your opinion of the magazine.' This proved sufficient for the purpose: only adult literacy learners were assessed, and only a very quick test was needed.
If only the task paraphrased above is used, we suggest allowing up to 15 minutes.

This would need to be piloted, and we estimate that the cost would not exceed **£30,000**

Reading

For the scoping study we recommend that the version of the 2002/03 reading and numeracy tests that John Bynner and Sam Parsons modified for the BCS70 survey in 2004 are evaluated using confirmatory factor analysis.

Guideline cost: £30,000.

5. NUMERACY: ADDITIONAL RECOMMENDATIONS

A special case

Numeracy represents a special case: we are recommending that *two* alternative sets of items are subject to confirmatory factor analysis in the pre-pilot test.

Numeracy colleagues have expressed some unease about the 2004 numeracy instrument, one of the instruments we nevertheless recommend including in a pre-pilot test. The 2004 test has itself been extensively piloted and provides for direct comparability with the items used in the 2003 Survey. However, so as not to limit the options available to DIUS, we are recommending that we include in a pre-pilot test a second and alternative set of numeracy items. In this section we focus on this second option, and specify how those alternative items would be arrived at.

Items would be identified following a three staged process:

1. Identify facility ratings of numeracy items used in the 2003 SfL Survey, with reference to Gillespie, J. (2003).[1]
2. Determine an appropriate alignment measure: that is, the degree of alignment, expressed as a %, to the level to which the item was assigned.
3. Identify items that perform well on that criterion, and assess these with reference to the ALL Numeracy Team's scheme of five 'complexity' factors.[2] This should be undertaken by at least two people, working independently, who compare assessments and arrive at an agreed set of items to be tested in the pre-piloting phase.

   Estimated time and cost: one day - £700.

Complexity

[1] *The National Basic Skills Survey of Adults in England 2002-3. Summary Numeracy Report - 7 08 2003*. Nottingham: University of Nottingham, School of Education, CDELL.

[2] See: Gal, I. (2007). *Assessment of Adult Numeracy in PIAAC: A Conceptual and Development Framework. 1-2 November 2007, Paris* (No. COM/DELSA/EDU/RD(2007)3). Paris: OECD; Gal, I., van Groenestijn, M., Manly, M., Schmitt, M. J., & Tout, D. (2003). Numeracy in the Adult Literacy and Lifeskills (ALL) survey: An overview and sample items; Gal, I., van Groenestijn, M., Manly, M., Schmitt, M. J., & Tout, D. (2005). Adult numeracy and its assessment in the ALL survey: A conceptual framework and pilot results. In T. S. Murray, Y. Clermont & M. Binkley (Eds.), *Measuring Adult Literacy and Life Skills: New frameworks for assessment* (pp. 137-191). Ottawa: Statistics Canada.

On this alternative, all SfL test items should be screened using the ALL Numeracy Team's scheme of five 'complexity factors', in order to ensure that items are correctly assigned to levels.

Apparently similar questions may be very different in their levels of difficulty. The ALL Numeracy Team has developed a scheme of five 'complexity factors' to account for this. The scheme is designed to enable both an explanation of observed performance in terms of underlying cognitive factors and the development of a complexity-rating scheme used to guide the construction of assessment tasks. These complexity factors are:

Textual aspects

1. Type of match/problem transparency: *obvious/explicit* - to - *embedded/hidden*

2. Plausibility of distractors: *no distractors/information all there* - to - *several distractors/information not all there*

Mathematical aspects

3. Complexity of mathematical information/data: *Concrete/simple* - to - *abstract/complex*

4. Type of operation/skill: *simple* - to - *complex*

5. Expected number of operations: *one* - to - *many*[3]

Additional recommendations

The alternative test should include a significant number of contextualized questions

Test items should be presented primarily through visual and oral media to make them as accessible as possible to people with literacy or language difficulties. Assuming laptop delivery, as in the 2003 Survey, this entails using written and spoken words or numbers, as well as using pictures and photographs.

Test items should include open and closed, long and short questions.

The test should include both multiple choice and short answer questions.

Questions or tasks should include some tasks where the required information is supplied and other tasks where it is implied, assumed or buried in irrelevant material.

---

[3] Manly, M., Tout, D., van Groenestijn, M., & Clermont, Y. (2001). What makes one numeracy task more difficult than another? In M. J. Schmitt & K. Safford-Ramus (Eds.), *A Conversation between Researchers and Practitioners. Adults Learning Mathematics - 7. Proceedings of ALM-7 the Seventh International Conference of Adults Learning Mathematics - A Research Forum* (pp. 78-85). Cambridge, MA: National Center for the Study of Adult Learning and Literacy (NCSALL), Harvard University Graduate School of Education, in association with Adults Learning Mathematics - A Research Forum (ALM); Gal, I., van Groenestijn, M., Manly, M., Schmitt, M. J., & Tout, D. (2005). Adult numeracy and its assessment in the ALL survey: A conceptual framework and pilot results. In T. S. Murray, Y. Clermont & M. Binkley (Eds.), *Measuring Adult Literacy and Life Skills: New frameworks for assessment* (pp. 137-191). Ottawa: Statistics Canada.

The data collected through piloting the test should be subjected to confirmatory factor analysis to evaluate check its dimensionality against that of the SfL, BCS70 and Army equivalent tests

Estimated time and cost: **30 days for each of 3 people; £30,000**

6. INFORMATION AND COMMUNICATION TECHNOLOGY

There is extensive scope for development of the ICT SfL assessment instruments to enable them to better measure what they set out to measure. What follows are roughly costed proposals for developments that would maintain a good degree of comparability with the original survey, whilst also allowing for comparisons with tests that are likely to be useful in the future.

We divide this into two sections the first of which brings the instruments into better alignment with the SfL standards, and the second of which would bring the instruments into better alignment with emerging forms of testing and ICT use.

Confirmatory factor analysis is again needed to evaluate the dimensionality of this new test as part of the development work we are proposing.

Refinement of present instruments to better match ICT SfL Standards

<u>1a) Differentiation between levels</u>

- Extend the Practical Assessment to differentiate between Level 1 and Level 2
- Extend both the Awareness and Practical Assessments to differentiate the Entry level classification into at least two stages, Entry 1+2 and Entry 3.

<u>1b) Content</u>

Incorporate not only purely technical questions but also issues such as: ‘following safely and security practices’, ‘select and use a variety of appropriate sources of information’ and’ present information in ways that are fit for purpose and audience’ (all statements drawn from the Level 1 standards)

Examples drawn from the instruments that we developed for the ICT Effective Practice project and that could be developed for the Awareness Assessment include:

24) You get an unexpected email from a friend with the message "Best wishes, your friend" and file attachment **postcard.zip**
What is the safest option?

- **A** ❒ Open the attachment.
- **B** ❒ Email your friend to confirm they sent you an attachment before opening it.
- **C** ❒ Save the attachment, and then open it using a file compression program.
- **D** ❒ Forward the email to friends.

32) You are a customer of 'YourBank" and receive the following email:

£ YourBank

Dear Sir/Madam,

We have recently changed the servers and cannot verify the information we have for you. We nead to update the information we hold for you as this affects your ability to log on and use the internet banking service.

Please use the folowing link to update your account information.
https://www.yourbank.co.uk/ibank/Request?CmdId=DisplayLoginPage

Yours sincerely,

James Vendeer
YourBank Internet Customer Services

Do you?

- **A** ❒ Follow the link and confirm your details.
- **B** ❒ Ring customer services to confirm this has happened.
- **C** ❒ Reply to the email and include your details.
- **D** ❒ Reply to the email telling customer services what you think of them!

1c) Methods and costs

Development of test items, testing with a sample of 100, and consequent further development of test items.

Three months work for three people. **Total costs £100,000**

Significant development of instruments to better measure ICT skills

2a) Develop to capture concept of 'purposeful use'

In order to capture elements of 'purposeful use' (which the ICT SfL standards recommend) the Practical Assessments could be extended to use more realistic scenarios. The following two examples are taken from the iSkills test from ETS -

now directed at College entrants, though initial thinking about this test stemmed

Scenario: Managing Files *Cleaning a Computer* You are giving your old computer to your younger brother and want to delete any personal emails. You also want to save emails related to your job search so that you can transfer them to your new computer. **Task:** • Read the emails (double-click on any that you want to view) and drag them into the appropriate folders. • After you have finished, delete the emails you do not want to save.	**Scenario: Creating a Slide** *International Student Club* You are president of the International Student Club, which shares meeting space with the Drummers' Club. Because both clubs have good reasons for using the space on Monday, Tuesday and Thursday afternoons, your club meetings have become crowded and noisy. Your club has asked you to talk with the faculty advisor about getting a new room for your meetings. The club secretary has sent you an email containing suggestions from different club members, and now you need to prepare a slide that makes your case for a new meeting room in a clear and persuasive set of bullets. **Task:** • Using the email as the source for your points, create a single persuasive slide to use in your discussion with the faculty advisor. • Format your slide appropriately, using the fonts and text styles provided in the slide creator. • To sequence bulleted text in a slide, select the bulleted text that you want to reorder and use the up or down buttons to move it. To delete a bullet point, select the bulleted text that you want to delete and click on "Remove text."

from a focus on basic information literacy needs. Similar software should be developed to allow interactive testing of ICT SfL.

2b. Incorporating test items related to the changing world of ICT.

Elements related to the changing world of ICT need to be incorporated, and whilst this would affect compatibility with the previous test, failure to incorporate some of them would be likely to reduce the scope for comparability with future assessment instruments. Two areas in which items might be developed are: e-Government and data security

2c) Methods and costs

Software development, development of test items and trialling:
£100,000

7. PILOTING AND USING THE 2004 LITERACY/NUMERACY ASSESSMENTS

The BCS70 2004 assessments were extensively developed and piloted. We know the length of time they take to administer. We also know that most respondents were able to complete them administered by interviewers using a computer in their own homes without difficulty.

We are assuming that the piloting of the test as part of the total interview would be a job for the agency appointed to undertake the survey. In 2003 this was BMRB (NRDC used NatCen for the 2004 Survey.)

The assessments are programmed in Blaise, the CAPI programme used by NatCen. For the ongoing longitudinal army study, the 2004 assessments need to be extracted from the main 2004 survey and attached to the 2008 SfL questionnaire.

To run the assessments (and accompanying questionnaire) the computer/interviewer needs to be covered by a Blaise licence. NatCen has a Blaise licence and all employees/interviewers are covered by this licence.

To carry out a pilot using NatCen interviewers, NatCen would have to be recruited as soon as possible - they have limited fieldwork capacity this year. In recent discussions in connection with the army study, they have indicated some limited capacity in August but then not again until early 2009.

It would not be viable to carry out a pilot study using NRDC personnel: the ‘sample’ is the general public, not a selected army sample found in one location. A Blaise licence would also need to be bought. **We do not know the cost of this, but we understand that it is prohibitive**. The fee was waived for Sam Parsons to carry out the army pilot, as the fieldwork extended over a few days only, and NatCen did not have fieldwork capacity at the time. Sam Parsons had to seek permission directly from Blaise and to sign a licence agreement document.

A new questionnaire

This would have to be developed and programmed into the CAPI. The pilot would be used to test the length of the questionnaire, to ensure that it is within the proposed 1.5 hours.

If NatCen were not going to carry out the SfL survey (either because they opted not to tender or because they failed to win the contract) then additional costs and time would need to be considered for adapting the Blaise programme to the language of the CAPI programme used by the successful fieldwork agency.

8. SUMMARY OF KEY RECOMMENDATIONS

Main recommendation: a pre-pilot evaluation

- Undertake a confirmatory factor analysis to evaluate the measurement properties of the literacy, numeracy and ICT items, and the dimensionality of the tests as a whole across three different target populations or survey research settings.

Literacy

- Not to devise separate tests or items for spiky profiles.
- Use the writing assessment within the set of Go! tools developed by NFER for NRDC.
- Evaluate the modified version of the 2002/03 reading test as used in the BCS70 survey in 2004.

Numeracy: 2004 test

- Evaluate the modified version of the 2002/03 numeracy test as used in the BCS70 survey in 2004

Numeracy: alternative test

- Evaluate a second numeracy instrument.
- Instrument to be based on selected items from the SfL numeracy instrument
- Items to be evaluated using the ALL scheme of five ‘complexity factors’.
- Include a significant number of contextualized questions.
- Include items presented primarily through visual and oral media, so as to make them accessible to people with literacy or language difficulties.
- Include open and closed, long and short questions.
- Include both multiple choice and short answer questions.
- Include some tasks where the required information is supplied and other tasks where it is implied, assumed or buried in irrelevant material.

ICT

- Extend Practical Assessment to differentiate between Level 1 and Level 2
- Extend both the Awareness and Practical Assessments to differentiate the Entry level classification into at least two stages, Entry 1+2 and Entry 3.
- In order to capture elements of ‘purposeful use’ the Practical Assessments could be extended to use more realistic scenarios.
- Elements related to the changing world of ICT need to be incorporated.