

For Official Use

COM/DELSA/EDU(2004)9



Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development

02-Nov-2004

English - Or. English

**DIRECTORATE FOR EMPLOYMENT, LABOUR AND SOCIAL AFFAIRS
DIRECTORATE FOR EDUCATION**

**COM/DELSA/EDU(2004)9
For Official Use**

PIAAC DRAFT STRATEGY PAPER

Policy Objectives, Strategic Options and Cost Implications

15-16 November 2004 in Stockholm

Contact: Alistair Nolan, Indicators and Analysis Division, Directorate for Education [Tel: +33 1 4524 1386; Fax: +33 1 4430 6177] or Koji Miyamoto, Employment Analysis and Policy Division, Directorate for Employment, Labour and Social Affairs [Tel: +33 1 4524 8554; Fax: +33 1 4524 9098]

JT00172946

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format

English - Or. English

**PROGRAMME FOR THE INTERNATIONAL ASSESSMENT OF ADULT
COMPETENCIES (PIAAC)**

DRAFT STRATEGY PAPER[†]

Policy Objectives, Strategic Options and Cost Implications

OECD, Directorate for Education (EDU) and
Directorate for Employment, Labour and Social Affairs (ELSA)

[†] This paper reflects expert opinion from several countries, partly obtained through seminars organised by national authorities in Canada, Germany, Japan, Sweden, the United Kingdom and the United States. Jon Hales (National Centre for Social Research, UK) assisted the Secretariat with the assessment of methodological, operational and resource implications of the draft strategy.

TABLE OF CONTENTS

INTRODUCTION AND SUMMARY	5
PART A. THE CONTEXT TO PIAAC	8
1. The policy relevance of a new assessment of adult competencies.....	8
2. The advantages of assessing adult competencies internationally.....	12
3. Synergies with IALS and ALL	12
PART B. DEFINING STRATEGY CHOICES	13
4. The need to define strategy choices	13
5. Which age groups should be covered in a first PIAAC assessment cycle ?	13
5.1 Merits/demerits of a first-cycle focus on the youth cohort.....	14
5.2 Merits/demerits of a first-cycle focus on a cohort of older adults.....	15
5.3 Merits/demerits of a first cycle covering all age groups.....	15
5.4 Merits/demerits of a first cycle covering all age groups but with age-specific over-sampling .	16
6. Competency domains and methods of assessment.....	17
6.1 Assessing generic skills using a job requirement approach.....	17
6.2 Obtaining information relevant to the employers’ perspective	18
6.3 Testing the generic skills of the adult population.....	22
6.4 Other information to be collected on respondents.....	29
6.5 Social capital and citizenship	30
6.6 The Number of Skills Domains to be Assessed.....	31
6.7 Preparing Complementary Qualitative Studies	32
7. Links to subsequent assessment cycles	32
7.1 Merits and demerits of a longitudinal survey	33
PART C. A FIRST CONSIDERATION OF RESOURCE REQUIREMENTS	36
8. Approximate cost implications of different strategy options	36
8.1 Factors likely to cause cross-country variation in costs	39
8.2 Funding for the preparatory phases	43
9. Management and governance options.....	43
PART D. THE SECRETARIAT’S STRATEGY PROPOSAL.....	44
REFERENCES	48
APPENDIX A. EXISTING INTERNATIONAL COMPARATIVE ASSESSMENTS OF SKILLS.....	51
A.1 International Adult Literacy Survey (IALS).....	51
A.2 Adult Literacy and Life Skills Survey (ALL).....	52
APPENDIX B. ADAPTING THE JOB-REQUIREMENT APPROACH FOR AN INTERNATIONAL SURVEY	53
B.1 Proposed procedure.....	55
B.2 Resource implications	57

B.3 Proposed schedules58

APPENDIX C. CLASSIFICATION OF SURVEY COSTS59

APPENDIX D. IMPLICATIONS OF USING A COMPUTER TO ADMINISTER SURVEYS63

APPENDIX E. AN OUTLINE OF INITIAL COST ESTIMATES AND RELATED ASSUMPTIONS...64

E1. Assumptions that are common to all options.....64

E2. Assumptions specific to entire age spectrum option65

E3. Assumptions specific to the option of focusing on youth (aged 18-35) and the option focusing on older adults.....65

E4. Cost estimate for survey of 4,000 adults aged 18 to 65.....66

E5. Cost estimate for survey of 4,000 adults aged 18 to 35 (with identical costs assumed for the age group 45-65).....67

E6. Cost estimate for survey of 4,000 adults aged 18 to 65, with 100% extra 18 to 3568

INTRODUCTION AND SUMMARY

1. The preparation of this paper follows the deliberations at the first meeting (26-27 April 2004) of the International Expert Group (IEG) to consider the development of an OECD strategy for a Programme for the International Assessment of Adult Competencies (PIAAC). The paper aims to: a) describe the policy relevance of a new assessment of adult competencies; b) formulate strategic options related to the choice of age groups and competency domains to assess; and c) identify cost implications of the different options.
2. Part A outlines the background to PIAAC, describing key current policy concerns that directly relate to adult competencies. These concerns include such issues as: linkages between productivity, the demand for competencies and globalisation; the skills implications of rapid population ageing; and problems of labour market transition among young adults. An overview is also presented of a broad spectrum of themes on which PIAAC could generate new policy insights, ranging from the performance of education systems and labour markets, to a series of wider economic and social concerns.
3. Part B addresses the need to make strategy choices and the definition of those choices. For a variety of reasons, it is critical to initiate the conceptualisation and design of PIAAC based on clearly defined policy goals and priorities. Doing so will provide the necessary focus for a strategic appraisal of the competency domains, age cohorts and policy concerns that PIAAC could feasibly address. It is also noted that a strategic view of PIAAC is essential because suitable measurement technologies do not currently exist for all competency domains that are of policy interest. Consequently, strategy choices are required in order to specify the competencies that should be assessed in a first assessment cycle, as well as those that are of policy interest but that require the development of measurement instruments so that they might be assessed in the future.
4. To facilitate strategy choices, Part B includes a detailed consideration of the merits and demerits of four possible age-group foci for a first PIAAC assessment cycle. These four options are: a focus on the youth cohort (possibly 18-35 years of age); a cohort of older adults (possibly 45 to retirement age); a focus on all age groups; and a focus on all age groups but with age-specific over-sampling. With a view to facilitating choices among the competencies that might be assessed, Part B also provides a synopsis of the literature on how different competency domains affect individual and aggregate outcomes. The competency domains considered are in: literacy (including numeracy and health literacy); new information and communications technologies; team-working; communication; problem solving; and non-cognitive skills. The availability of assessment methods for international comparative purposes is described for all of these domains. Part B further considers the prospect of assessing work-based competencies in an indirect way, through a so-called job-requirement approach.
5. A number of countries have expressed strong interest in including an employers' perspective in PIAAC. Doing so could create unprecedented international data on a variety of important themes linked to the demand for skills, the nature of skills shortages, enterprise-level strategies to develop skills, causes and consequences of technological change, and many other issues. Responding to this interest, Part B outlines how the employers' perspective could best be assessed. Attention is likewise given to the advantages and

drawbacks of longitudinal survey design, as well as the possibility of conducting optional qualitative studies that might become a complementary component of PIAAC.

6. Part C provides a first outline of possible resource requirements for PIAAC. Cost estimates for each of four potential age-related assessment options are sketched. The estimates involve assumptions that are explicitly stated in the main body of the text and in Appendix D. The drivers of likely cost variations across countries are also described.

7. In seeking to bring greater focus to the strategy decisions that developing PIAAC requires, the text is interspersed with questions for IEG members. In summary, the key strategy questions that require agreement by the IEG are:

1. *Which age group(s) should be the focus of a first PIAAC cycle and for what policy purposes? If it is agreed to sample all age groups with age specific over-sampling, which age group should be over-sampled in a first cycle ?*
2. *Should PIAAC include a survey of employers? If so, should employee information be gathered starting from a sample of employers, or should a survey of employees be the basis for a selection of employers ?*
3. *Which competencies should be assessed in a first PIAAC cycle and why ? Which competencies should be assessed using the job requirement approach and which using other forms of assessment (i.e. direct testing), given that the non-employed would also need to be assessed ?*
4. *For which competencies would there be an interest in a programme of research aimed at the future development of measurement instruments ?*
5. *Which items of contextual information on interviewees are considered of greatest interest?*
6. *Do experts consider that complementary research could add value to a first assessment cycle (e.g. in the form of qualitative studies). If “yes”, which subject areas are of greatest interest ?*
7. *What are the views of the IEG regarding the Secretariat’s strategy proposal, and which of the two options (1 or 2, see below) is preferred ?*

8. Representatives from a number of countries have recently asked that the Secretariat propose a strategy for PIAAC from among the many possible permutations of age and policy foci. Part D outlines the Secretariat’s proposal. Essentially, under this proposal, ***PIAAC would be a multi-cycle programme of assessment covering, over time, all the main policy concerns articulated by the IEG.*** The first cycle of data collection is envisaged in 2008~9, with reporting in 2010. Subsequent cycles would be administered at five-year intervals.

9. The Secretariat proposes that ***each cycle cover all age groups including the non-employed with an age specific focus or over-sampling.*** Having data from across the age spectrum – even if only in benchmark form – is considered necessary in order to explore a number of important policy issues. Age-specific focus might be made on: young adults during the 2008~9 cycle; middle-aged adults during the 2013~14 cycle; and an older cohort of adults for the 2018~19 cycle. One of two possible options could be considered:

1. A specific age-group focus in each cycle. We propose to begin with youth, followed by mid-age adults, and then an older cohort (definition of the age range covered by each category can

be decided by the IEG). However, each cycle will also include a small sample of individuals from across the age spectrum.

2. A coverage of all age groups in each cycle with over-sampling of youth in the first cycle, middle-age adults in the second cycle, and old-age adults in the third. This means that the sample size has to be large enough.

10. ***For the first cycle, PIAAC would directly assess competencies in the domains of literacy, numeracy, problem-solving and ICT.*** Subject to validation of the use of the job-requirement approach in an international cross-cultural setting - ***a survey of job requirements would be carried out simultaneously, again across all age groups.*** The combination of the job-requirement approach with direct assessment would be a feature that distinguishes PIAAC from other international assessments of skills, and would allow the examination, in new ways, of a variety of policy issues. To permit the assessment of additional competency domains, and based on priorities expressed by member countries, research could also begin, at an early date, with a view to fielding new instruments in the second and subsequent cycles of PIAAC.

11. ***PIAAC would include an enterprise survey,*** which could shed light on a broad array of policy issues not addressed in previous international skills assessments. The Secretariat proposes that feasibility work begin now to determine the viability and *modus operandi* of an enterprise survey.

PART A. THE CONTEXT TO PIAAC

1. The policy relevance of a new assessment of adult competencies

12. There is consensus that competencies (understood as the combination of knowledge, skills and attitudes) are essential to success in a wide range of key outcomes for individuals, firms and society. This section briefly describes two reasons why adult competencies occupy an especially prominent position in today's policy foreground: the skill-intensity bias in labour demand associated with new technologies and globalisation, and the effects of rapid population aging.

13. There is evidence that greater competencies are needed to take advantage of new technologies and adapt to the challenges of "globalisation". Evidence of growing wage gaps between more and less-skilled individuals has drawn attention to the emergence of a skill-intensity bias in labour demand in OECD economies. Various explanations have been advanced to account for the rising employment and wage premia that skills attract. These include: autonomous technological progress that may have benefited skilled more than unskilled labour; the effects of increasing international integration in markets for labour, goods and services; and organisational change, such as the greater devolution of responsibility and management to employees (Mariachristina *et al*, 2003). Furthermore, many observers hold that changes in workplace organisation have led to shifts in the demand for different *types* of skills. It is often claimed that the new workplace requires workers with multiple skills, particularly general skills in such areas as problem solving and communication, as well as non-cognitive skills such as interpersonal competencies and the ability to work autonomously. In the face of evidence for an increasing role for skills in advanced economies, many OECD governments are concerned that a "skills gap" could threaten their future growth prospects.¹

14. The rapid aging of OECD populations also raises critical skills-related concerns. While life expectancy is rising, falling birth rates mean that Europe's workforce will begin to contract from 2010 onwards. At current employment rates, the number of workers in the 15 countries that were members of the European Union prior to the 1st of May 2004 could fall by 14 million over the next 25 years. Were this projected contraction to occur, the region's economic output could decline by around 7 percent, absent a major increase in immigration.² Growing fiscal burdens associated with rising payments for pensions and health care – and high levels of economic inactivity in an expanding cohort of older citizens - will require

¹ An analysis by the OECD Secretariat of the causes of economic growth showed that rising labour productivity accounted for at least half of GDP per capita growth in most OECD countries over the period 1990-2000 (OECD, 2003). With educational attainment being the most frequently used proxy for human capital, available studies based on neo-classical models suggest that a one-year increase in average education raises the level of output per capita by between 3 and 6%. Studies of the "new-growth" variety – while not uncontested - find that the same increase in average education raises the rate of growth by around 1% (Sianesi and Van Reenen, 2003; De la Fuente and Ciccone, 2003).

² "Europe's costly resistance to foreign workers", International Herald Tribune, Friday July 23, 2004, page 2.

that workers remain economically productive later into life.³ Accordingly, older workers will need to possess the competencies required for longer careers. Governments thus need to have empirical knowledge of the competencies of older workers. The phenomenon of population ageing also highlights the increasing importance of efficient transitions to productive employment among young people. This is a challenging task: youth unemployment has increased substantially in recent years, reaching an OECD average in 2003 of 13.6% among youth between the ages of 15 and 24 (OECD, 2004)⁴. Such high, and in some countries persistent, levels of youth unemployment raise important questions about possible mismatches between the skills possessed by young people and the skills required by employers.⁵

15. PIAAC could generate new insights into the nature of these overarching policy challenges. In the most generic terms, PIAAC would provide internationally comparable data and analysis on: a) the types of competencies needed in today's economy and society, recent changes in competency requirements, and the extent to which the competencies possessed by adults are effectively utilised in the labour market, b) the type of education, training and labour market policies that help to improve the formation of key competencies. In particular terms, PIAAC could assist the formulation of policies that improve labour market performance, education system performance and general socio-economic performance. The following subsections describe a selection of issues in these policy fields on which PIAAC could provide new data and analyses.

Labour market performance

16. PIAAC could improve understanding of:

- How the competency requirements of work are changing nationally and internationally over time (and how firms are responding); the extent and types of competency mismatches and the effectiveness of policy settings relevant to addressing this problem.
- The extent to which competencies are effectively utilised in the labour market, shedding light on how and why the rate of utilisation of competencies differs across countries (for instance, a measure of the extent to which supplied generic skills are effectively utilised could be obtained by comparing the results of tests of generic skills with those from the job-requirement method).
- How the use of competencies varies by age, gender, and occupation.
- How adult competencies relate to the probability of non-employment and the effectiveness of policies in addressing the transition to work of both unemployed and inactive adults. How different competencies relate to job mobility and other labour market outcomes including early retirement decisions.

³ The costs for society of economic inactivity among older adults can be significant. For instance, a recent report from the UK estimated that the relatively lower level of employment among older workers costs the United Kingdom economy between £19-£31 billion a year in lost output and taxes and increased welfare payments (National Audit Office, 2004).

⁴ This figure is especially high in Poland (43.0%), Italy (26.3%) and Greece (25.1%).

⁵ For instance, policy-makers need to know whether skills mismatches are attributable to school systems that produce young adults without the level or type of necessary skills. Similarly, it is essential to know whether skills mismatches reflect deficient enterprise-level training aimed at developing specific skills in young adults who enter the labour market possessing only generic skills.

- The efficacy of different adult learning arrangements in producing competencies (e.g. by types of provider; whether full-time or part-time; whether classroom-based or through on-the-job training; using national qualification systems; with different financing mechanisms [such as various tax incentive schemes, legal obligations to finance training, and loans]; or with other incentives, such as day-care facilities). Insight could also be gained on the short- and long-run effects of different types of training programme.
- How the acquisition and loss of competencies are affected by labour market background (e.g. by work experience, type of contract, duration of non-employment, earnings, participation in training schemes, availability of social benefits, etc.). Moreover, it is pertinent to know whether skills are eroded with age, or become obsolete vis-à-vis employer job requirements, and whether participation in life-long learning helps reduce skill obsolescence and upgrade skills of older adults.
- The competencies possessed by immigrant populations.
- How “place” affects the formation of competencies and the performance of education systems. For instance, in the United States, workers in cities earn up to 33% more than non-urban counterparts (Glaeser and Maré, 2001). These wage differences are not fully explained by differences in ability. Cities appear to accelerate the accumulation of human capital. But how this occurs – for instance through more efficient labour markets or faster learning - is unclear.

Education system performance

17. PIAAC could improve understanding of:

- The competencies that are key to effective school-work transitions and how these competencies are best produced through initial education (e.g. through: work-oriented learning in schools; different work/study combinations; development of the vocational education sector; the involvement of employers in the design of occupational qualifications; services designed for individual needs; and the availability and mode of provision of information and guidance systems).
- The entire process of competencies acquisition and loss following initial education, and for a variety of competency types. Evidence could be gathered on the relative importance of initial, work-based and informal learning in the production of competencies used in the workplace (and at different stages of the life-cycle). Among other possible uses, this could indicate which areas of competency formation should be a focus for which types of institution (schools, colleges, universities, work-based training etc.) and what might be the appropriate breadth in any school curriculum.
- The rate of return to particular competencies (there are, of course, many studies on the returns to education generally).
- Whether there are different types of education that are more effective in providing a basis for the subsequent development of competencies.
- How different adult learning arrangements impact on equity in the acquisition of competencies (e.g. by level of income, occupation, gender, immigrant or non-immigrant status, minority or non-minority grouping).

- A possible international comparative framework within which national qualifications from both academic (and perhaps vocational programmes) could be situated.
- How national education systems that differ in terms of funding modality compare in terms of the level and distribution of competencies produced ?
- The proportions of people who are educated above the level required at work. Greater understanding could emerge of issues around the optimal proportion of young people staying on in full-time education, and whether over-education involves resource waste over the long-run.

Other aspects of economic and social performance

18. PIAAC could improve understanding of:

- The non-market benefits – such as civic and political participation, and social cohesion - associated with different competencies.
- The distribution of competencies among different social groups, its causes and consequences.
- How different competencies directly and indirectly affect a wide variety of health outcomes. In addition, assessment of health literacy could indicate mismatches between critical health documentation and the reading levels of the adult population, particularly among at-risk groups.
- How competencies vary by gender, and how significant gender differences affect individual and societal outcomes. For instance Coulombe *et al*, (2004) show that – for a variety of possible reasons - the strength of the relationship between literacy and macroeconomic performance is considerably greater for women’s rather than men’s literacy skills. And issues relating to gender, competencies, and levels of remuneration could be investigated in a variety of ways.
- Quantified links between human capital and macroeconomic performance. For instance, measurement of how skills affect labour productivity is hindered by insufficient variance in key parameters (Heckman and Vytlačil, 2000). This problem could be addressed through an assessment that covers a sufficiently large number of individuals. Similarly, average educational attainment is the usual proxy for human capital in analyses of macro-economic outcomes. Educational attainment is however a poor proxy of human capital, as many individuals possess fewer or more skills than indicated by their formal education. PIAAC could permit more accurate measurement of the stocks of human capital than is provided by indicators of educational attainment. By identifying the relative importance to growth of (different) competencies – across countries with a wide range of economic and social conditions - PIAAC could highlight where the greatest growth returns are to be had for different education investment strategies (among many other issues that PIAAC could address in this field).⁶

⁶ For example, Coulombe *et al* (2004) found that across fourteen OECD countries the proportion of adults with the highest levels of literacy skills has little impact on differences in GDP per capita. Such a finding calls, *inter alia*, for further investigation of the skills possessed by high-level knowledge workers but not captured in measures of literacy.

2. The advantages of assessing adult competencies internationally

19. There are a number of advantages to undertaking an assessment of adult competencies internationally rather than through national assessments alone. Such advantages were outlined at the first meeting of the IEG. For instance, because cross-country variation in a range of policies and institutional settings is greater than intra-country variation, an international assessment could in principle provide more policy-relevant data and analysis than a compilation of purely national assessments (if the latter were not strictly comparable). Furthermore, and most obviously, assessing cross-country differences in the level and distribution of competencies - and relating these to a broad spectrum of economic, social and policy conditions - would help policy makers to assess the comparative strengths and weaknesses in their country's past, current and prospective achievements. In addition, benefits would derive from exploiting synergies and economies of scale that are generated through international collaboration in the development of new assessment instruments and the interpretation of data. For instance, all countries could draw on institutional capacities and particular fields of expertise in other participating countries.

3. Synergies with IALS and ALL

20. It is essential that any new assessment builds upon the experience of existing surveys, notably the International Adult Literacy Survey (IALS) and the Adult Literacy and Lifeskills Survey (ALL). Appendix A provides a detailed description of IALS and ALL. A new international assessment could take IALS and ALL further in a variety of ways, since PIAAC is likely to include elements from these surveys. In particular, PIAAC could identify the relationship between the competencies that people from various socio-economic groups possess and those that are used at work. PIAAC could enlarge the coverage of countries, possibly adding to analytic power in the examination of some key policy questions. PIAAC could likewise include tests for domains not previously assessed under IALS or ALL, such as planning and high-level communication skills. The assessment of competences such as problem-solving and teamwork might also be approached in new ways, possibly by using the job requirement approach discussed below. Finally, there would be a potentially valuable chronological synergy with ALL. This is because the second round of data collection under ALL is set to occur between 2004 and 2006, with a first round of data collection under PIAAC possibly taking place in 2008/9.

PART B. DEFINING STRATEGY CHOICES

4. The need to define strategy choices

21. At the first IEG meeting countries agreed that it would not be possible to address all possible policy concerns facing all countries in a single survey, and that it would be advantageous to spread the multiple policy concerns over a number of survey cycles. For this and other reasons, it is essential that strategy choices be made in the design of PIAAC. Policy objectives, age foci and competency domains should be specified for a first (and possibly subsequent) assessment cycles. Clearly, a danger exists in attempting to tackle too many policy concerns in a single cycle. Overloading a first cycle could entail a loss of analytic focus, with data spread too thinly to provide valid insights on critical policy concerns. Addressing an extremely broad array of competency domains and policy issues may also not be feasible for reasons of cost. Furthermore, if unrealistic goals are set early on, a subsequent realisation of cost constraints might lead to a budget-driven reduction in PIAAC's scope, possibly compromising policy relevance or scientific method. Therefore, it is critical to initiate the design of PIAAC based on a detailed statement of policy priorities. A clear expression of policy priorities is a necessary basis for a strategic appraisal of the competency domains, age foci and policy concerns that PIAAC could feasibly encompass.

22. Another reason why strategy choices are essential is because appropriate measurement technology is not currently available for all the competency domains that are of potential interest. Accordingly, while commencing assessment with proven measurement instruments, PIAAC should also include a programme of research aimed at developing measurement instruments for competencies that are of common policy interest. Developing new instruments is an expensive undertaking, and often beyond the financial reach of individual national institutions. However, such developmental work is a natural candidate for international cost-sharing. Such collaboration could reasonably be expected to result in affordable per-country costs. A strategic view of PIAAC is therefore needed, one that entails the specification of competency domains, age foci and policy issues to cover in a first assessment cycle, along with the identification of issues of interest that require instrument development in future. In this connection, it is relevant to note the equivalent experience with the PISA survey, in which instruments for international comparative assessment of various skills (such as problem solving) were not available for the first assessment cycle, but were developed and administered in subsequent assessment rounds.

5. Which age groups should be covered in a first PIAAC assessment cycle ?

23. This section outlines, in generic terms, the merits and demerits associated with four possible choices for the age-group focus of a first assessment cycle. These options are: a focus on the youth cohort; on an older age cohort; on all age groups; and, on the entire age spectrum, with over-sampling of a selected age cohort.

5.1 *Merits/demerits of a first-cycle focus on the youth cohort*

Merits

- In the light of the importance in many OECD countries of economic and social issues affecting the youth population (problems of youth non-employment, school-to-work transition issues, etc.) a focus on youth will ensure early policy relevance for many participating countries. Of particular policy imminence might be to examine the extent to which skills learnt at school are effectively used in the early years after entry to the labour market.
- Unravelling the distinct influences of the many factors that contribute to skills in older age cohorts can best be done by following the evolution of skills acquisition/loss over time, beginning at an early age. Focussing on a youth cohort in a first wave of assessment could facilitate the development of panel or synthetic-cohort data in subsequent assessment waves (see section 8.1 for a discussion of the merits of a longitudinal survey approach). In this connection, a focus on youth could also allow the development of synthetic-cohort data based on PISA findings (including, possibly, for some non-OECD economies).
- It would also allow PIAAC to address early on a relative paucity of data pertaining to youth. IALS compares the level and distribution of literacy skills among adults⁷, TIMMS examines 13 year olds, and PISA samples learning outcomes among 15 year olds⁸. International comparative survey data on young adults are less well developed. In this connection, there is also much that is not known about the skills possessed by youth, for instance with regard to their ICT, communication and interpersonal skills. And across OECD countries as a whole, more remains to be learned about the impact on school-to-work transitions of a range of factors outside the individual's control (such as a person's parents, their home background, characteristics of the places where they grew up, etc.).

Demerits

- It could entail a limitation of the scope for providing certain types of policy analysis. For instance, policy issues that require comparison of the competencies of young people with the competencies of individuals in other age groups would not be fully addressed. Similarly, policy issues that relate the competencies of the workforce to macro-economic performance could not be properly considered without measures of the competencies had across all age groups.
- Assuming that PIAAC takes the form of a household survey, a focus on any specific age cohort – whether youth or older workers - is likely to raise costs, given that a larger number of households need to be visited in order to identify individuals for interview (unless there was a way to identify individuals in advance).⁹

⁷ Appendix A provides further details on IALS.

⁸ There have been several assessments for school children conducted before PISA. They include the Six Subjects Study in the 1960s, the First and Second Mathematics and Science Studies in the 1970s and 1980s, the International Assessment of Educational Progress in the 1980s, and the Civics Education Study and the Trends in Mathematics and Science Study in the 1990s.

⁹ The preliminary estimates contained in Part C of this paper indicate that the cost of assessing a cohort of young adults (between the ages of 18 and 35) would be about 19% higher than the costs of assessing all adults (between the ages of 18 and 65). One of the assumptions underlying these estimates is that there is

5.2 *Merits/demerits of a first-cycle focus on a cohort of older adults*

Merits

- A key merit of focusing on a cohort of older adults in a first assessment cycle stems from the particular policy relevance of this cohort given the speed of workforce ageing over the next decade. As the first round of PIAAC might not be launched before 2008-2009 – and if the focus of a first assessment cycle were on youth - then a second cycle focusing on an older cohort might not take place until 2011-12 at the earliest. With rapid workforce ageing set to occur during the coming decade, conducting an assessment in 2011-12 focussed on an older cohort might miss an opportunity to provide timely policy-relevant data and analysis.
- As a corollary to the preceding point, the human capital embodied in older workers was acquired further back in time and is thus less relevant to today's labour market. Accordingly, information on skills gaps in this group could be of particular policy relevance.
- Depending on the age range of the older age cohort, a focus on this group could encompass prime-age workers who, both numerically and in terms of their skills, make a decisive contribution to national productivity. A first-cycle focus on this age cohort (depending on the selected age range) would thus permit the early production of policy relevant data and analysis for those countries that have expressed a principal concern with issues of skills, productivity and macro-economic performance.
- For older workers, because of lower rates of job turnover relative to younger individuals, information on occupational categories is more reliable and tells us more about the individual than is the case for younger workers. Data on occupational status might more reliably be used in a skills assessment, possibly avoiding some of the high costs of survey work.

Demerits

- A focus on older workers could entail a limitation of the scope for providing certain types of policy conclusion, particularly those that require comparison of the competencies of older people with the competencies of individuals in other age groups.
- Assuming that PIAAC takes the form of a household survey, a focus on any specific age cohort – whether older workers or youth - is likely to raise costs, given that a larger number of households need to be visited in order to identify individuals for interview (unless there was a way to identify individuals in advance).¹⁰

5.3 *Merits/demerits of a first cycle covering all age groups*

Merits

no means of identifying individuals by age in advance, an assumption that does not hold true for all countries.

¹⁰ See the observations in the preceding footnote. In part C of this paper, the estimated costs for a survey of older adults (between the ages of 45 and 65) are the same as for a survey of young adults (between the ages of 18 and 35).

- Various countries have articulated policy priorities that concentrate on either youth or older age cohorts. Adopting a broad age spectrum in a first cycle would allow PIAAC to address both areas of urgent policy priority and yield results relevant to all countries and to all age groups.

Demerits

- Since IALS and ALL cover all age groups, it is possible that there might be little value added over these surveys. Furthermore, an overly broad analytical focus resulting from an assessment over a wide age spectrum could possibly lead to policy conclusions the meaningfulness of which is limited.¹¹
- With given (and constrained) budgets, assessing all age groups could increase the severity of trade offs between the number of competency domains to be assessed and the analytic power of the data (as sample sizes may be fixed by budget limits). However, whether this is the case depends in part on choices regarding assessment methods. Adding further skills domains when assessing using the job-requirement approach may have only a small impact on survey costs.

5.4 Merits/demerits of a first cycle covering all age groups but with age-specific over-sampling

Merits

- This option would entail the merits described above in connection with the age foci choices of either youth or older workers. It would also entail some of the merits of the option of assessing the entire age range. In addition, it would involve fewer of the demerits of assessing all age ranges, as greater policy-relevant specificity could emerge in connection with the over-sampled cohort.

Demerits

- As noted in 5.3, an assessment covering all age groups needs to focus on competencies that are relevant to a broad range of contexts. It may be that this will reduce the scope of the common competency domains. For example, the contexts in which younger and older adults use new ICT may differ significantly, possibly making a single assessment of ICT competencies meaningless.
- Assessing across the age spectrum combined with age-specific over-sampling would raise costs. However, based on the preliminary estimates outlined in Appendix D, the costs that this option would entail would still be very similar to those required by the age foci on either youth or older adults.

Question for IEG members: in a first PIAAC assessment cycle, which one of the above age foci is to be chosen and why?

¹¹

For example, the contexts in which youths are using ICT may differ from those for older age groups. If this is the case, the use of a common instrument to measure ICT skills across both groups may fail to provide a meaningful assessment.

Question for IEG members: in a first PIAAC assessment cycle, if the all-age-groups-with-age-specific-oversampling option is preferred, which age group should be oversampled ?

6. Competency domains and methods of assessment

24. This section outlines options for a strategy that would assess a set of competencies central to policymakers' current and likely future concerns. In accordance with views expressed at the first meeting of the IEG, consideration is given to a strategy combining test-based assessment (of selected generic skills) with a job requirements assessment method. In addition, this section considers the possibility of obtaining information from employers, an option recently suggested to the Secretariat by a number of countries. Combined with a test-based assessment of employees, a survey of employers would have a number of attractive policy-relevant features, providing as it would information on the demand for skills as well as the skills that employees possess.

6.1 Assessing generic skills using a job requirement approach

25. The job requirement approach to the assessment of competencies was described in Green (2004) and the associated presentation at the first IEG meeting. The method relies on detailed self-reporting of the activities that a person's job entails. The job requirement method is based on some of the principles of job analysis used by occupational psychologists. The job requirement approach was used in the 1997 UK Skills Survey (England, Scotland and Wales) and a slightly modified and extended approach was used in the 2001 UK Skills Survey. A starting point here was that competencies had generally been estimated by indirect methods in the past. For example, substantial reliance had been placed on an individual's time in education and level of educational qualifications, even though these might have been acquired many years previously. Occupational classifications imply a level of competencies, but may fail to indicate whether an individual's competency level is below or above what is required for the job they are doing. In addition, the competencies needed for a given occupation probably change over time. More decisive in some ways was the fact that employers' decisions about hiring employees or assigning them to particular jobs often involved assessments of other aspects of competency, such as work experience, use of IT, communications or problem-solving, and none of these were being measured in a quantitative way.

26. In using the job requirement method it is assumed that the job's activities represent an unbiased measure of the job-holder's competencies. The method provides information on the competencies used at work, rather than the competencies people possess. Advantages of the method include the following:

- It allows the assessment of a wide range of competencies (Felstead *et al*, 2002, page 16). For instance, the UK Skills Surveys used measures of the following competencies: computing; problem-solving; checking; high-level communication; planning; horizontal communication; client communication; physical and technical know-how; literacy and numeracy;
- The competencies are measured in context;
- Competencies are measured based on activity reports made by the job-holder. The job-holder will be well-informed of the job's activities. S/he is also likely to be more accurate in describing job

activity requirements than in describing personal competencies and/or educational qualifications;¹²

- The approach is cost-effective.

27. Possible drawbacks of the job requirement approach are: there may be mismatches between competencies possessed and competencies used; there is subjectivity in the reporting of the job-holder, and the non-employed are excluded. However, problems of competency mismatching are less acute if one assumes, reasonably, that individuals would not continue to hold a job that involved activities requiring competencies that they did not possess. Problems of job-holder subjectivity can also be (partially) countered by cross-checking the job requirements reported by the job holder and the requirements of the same job as reported by a supervisor.

28. To date, the job requirement approach has been used with success in two surveys in the United Kingdom. A number of steps would be needed to adapt this approach for use in an international survey. At a minimum, these steps would be:

- To initiate the international piloting of this approach, develop a questionnaire that suits the purposes of PIAAC. To this end it is important to: (a) consider the type of generic skills to be covered; and (b) select questions that are relevant to international comparisons, in the light of potential cross-cultural biases.
- Engage a suitably qualified survey company in each of three countries, and meet to examine the English questionnaire.
- Translate the questionnaire (using double translation followed by reconciliation, that is, with three checks into the target language).
- Conduct face-to-face interviews in each of the three countries using the same questionnaire. However, to assess cross-country and cross-cultural validity an analysis in a cognitive laboratory in each country may be required (a mix of frequency and importance scales for measuring activities can also be tested during this pilot phase).
- Conduct a pilot survey of around 500 interviews, based on a probability sample, involving the development of the full repertoire of survey materials and their implementation.

A more detailed specification of the tasks required to adapt this approach for use in an international survey is contained in Appendix A.

6.2 *Obtaining information relevant to the employers' perspective*

29. A number of countries have indicated that PIAAC should also collect information on the employers' perspective on competencies. This could be achieved by conducting an enterprise survey.

¹² One should not under-estimate how much difficulty people may have with reporting the qualifications they hold. Older people may have acquired qualifications a long time ago and may not refer to them in day-to-day life. Furthermore, the names of qualifications change, so older individuals may struggle to find an equivalence between their qualifications and the current system. Even people who recently gained qualifications may need careful questioning to identify their qualifications with any accuracy. Another issue is whether the qualifications sought were completed in part or in full.

Critical information that could be extracted from enterprise surveys include: (a) characteristics of the firm; (b) firm-level activities in developing competencies among employees; and (c) the firm's demand for employees with different types and levels of competency. While information on some firm characteristics can, in principle, be obtained by asking employees (e.g. via surveys on individuals), doing so is likely to involve large measurement error. For example, key variables such as firm-size and capital stock, industrial classification and R&D activities are difficult for employees to specify. While employees would most likely know about *their own* training activities, they need not have a complete picture of the human resource development activities (including expenditures on training for different employees) across the firm. Firms' demand for competencies is also difficult for employees to assess. Thus, without conducting a firm-level survey, it would appear difficult to capture some of the essential information relevant to a number of policy concerns described in Part A of this paper.

30. Data on key background characteristics of enterprises that can be extracted from enterprise surveys include: (a) industrial classification, firm size, value-added, capital stock, and foreign capital shares; (b) other firm activities that relate to employee competencies, such as innovation, R&D expenditures and investment in and usage of ICT; and (c) the composition of employees by level of competency and/or task.

31. Critically, data on the background characteristics of enterprises could also include a profile of training practices, possibly taking account of information on expenditures on, and the duration of, each type of training, as well as the characteristics of employees who receive training.

32. The demand for competencies is perhaps one of the most important types of information that an enterprise survey could indicate, since this would help to identify competency mismatches. There are various direct measurements that may indicate excess demand for employees with particular competencies: (a) employers' perception of competency gaps, and (b) vacancy rates. There are also indirect ways through which competency mismatches can be captured, including: (a) changes in relative wages across occupation; (b) hours and intensity of work; (c) production levels; (d) flows of new entrants and leavers; (e) hiring standards; and (f) level of immigration by skill groups (note that [b] ~ [e] may indicate a firm's strategy to cope with competency shortages if it does not upgrade the competencies of its employees).

33. Although an enterprise-based survey could capture policy-relevant information, it would add significant costs if enterprise surveys are to be conducted on top of surveys of individuals (see Part C). However, there are ways in which both individual surveys and enterprise surveys can be done jointly in a cost effective manner. These are considered in the next section.

6.2.1 *Matching information from employers and employees*

34. If conducting enterprise surveys is considered to be a realistic option, linking individuals with enterprises will enrich the data substantially. This has not been done in past international skills assessments. However, there has been a recent development in labour market survey practice that matches employee and employer data. This is the so-called matched employer employee survey.

35. There are three ways in which a matched employer employee survey might be done. The first is to use independent individual and enterprise surveys and to link them by using a code, name or address that helps identify the firms that the individuals are working for. For example, for the United States, the Decennial Employer-Employee Dataset (DEED) matches the 1990 Decennial Long Form of the Census and the 1990 Standard Statistical Establishment List using address matching software (Hellerstein and Neumark, 2004). There are also data sets that have matched the National Longitudinal Survey of Youth (NLSY) with large-scale establishment data (Abowd and Finer, 1999). This matched survey has considerable analytical potential since the NLSY provides a large-scale longitudinal database containing a

rich set of variables.¹³ However, enterprise surveys that document detailed human resource development activities and the demand for competencies are quite rare in OECD countries. So new enterprise surveys would likely be needed.

36. The second way to obtain matched employer-employee data is by first identifying the firms (or use pre-existing enterprise surveys) and then identifying individuals (employees) by random selection of employees from a list of employees obtained from the employers. A number of the advantages and disadvantages associated with sampling individuals in the workplace are described in Box 1. Since firms can be identified using the national register of enterprises, and target individuals can be identified by asking employees, identification of samples may be relatively straightforward.¹⁴ This method was adopted in constructing the Workplace and Employee survey in Canada (Krebs *et al*, 1999), and the Workplace Employee Relations Survey (WERS) in 1998 and 2004 in the United Kingdom (Belfield and Wei, 2004). The Canadian data are particularly rich as they reflect a longitudinal survey which tracks enterprises and employees over a number of years. WERS 98 and WERS 2004 use a fixed number of employees (25 per establishment, or everyone in cases where there are fewer than 25 persons in the workforce). Respondents need to be selected from an up-to-date list, such as the payroll. The workforce may include individuals who are paid by the establishment but who usually work elsewhere, such as home-based employees. In 1998, the UK WERS involved a sample of just over 2,000 establishments with 10 employees or more in all sectors except agriculture. A sample of nearly 30,000 self-completion questionnaires was generated.

¹³ The data set was constructed by Abowd and Finer (2004) by linking the individual's job-related information provided by the NLSY79 (based on the 1979 cohorts) to information about firms that hired the same individual. They matched employer names, industry, employment size, location (reported in the NLSY79) with information provided by COMPUSTAT (Standard and Poors), information from the Center for Research on Security Prices at the University of Chicago, and the Directory of Corporate Affiliation (National Register Publishing Company). Altogether, about 4000 employers were matched with 6600 individuals for the period of 1986 and 1994 (for a total of 33 thousand observations).

¹⁴ It is possible to obtain employer data without contacting the organisation – from various databases. These data are often limited in scope and quality, and may have little or nothing to indicate the nature of the organisation's workforce or skills development activities.

Box 1. Conducting interviews in a workplace setting

In-home interviewing involves choosing a sample frame and selecting addresses and individuals. There is then a considerable effort on the part of interviewers to achieve high response rates. It may be worth asking whether this is the optimal approach, especially if there is to be a focus on the workforce currently in employment. An alternative is to implement PIAAC through a sample of employing establishments. This text box is intended to stimulate discussion of this approach, should members of the IEG feel this to merit more detailed consideration as an option.

There are a range of potential advantages and disadvantages to sampling from the workplace. Among the advantages are:

- Lower cost per interview, although gaining co-operation of employers can require considerable investment.
- The value of the interviewer's time may be maximised by one respondent doing the self-completion part of the interview while the interviewer is conducting the interviewer-administered part with another respondent,
- The ability to cross-check some aspects of each selected job with a manager at the establishment
- Because details of the establishment can be obtained from a manager - including information on training arrangements and use of IT at the workplace - respondents are only asked questions that focus on their competencies, which keeps the interviews shorter.

The disadvantages of sampling from the workplace include:

- A more limited scope to collect detailed information than in-home interviews, given that this approach involves a number of staff being absent from their work station and so may limit the length of each interview.
- Arguably, respondents might feel under pressure to return to work as quickly as possible, or feel inhibited about answering questions at their place of work; this could affect the quality of answers.
- There are two stages at which refusal can occur (the manager may refuse on behalf of the establishment and the individual may refuse), so it might be difficult to assess the extent of bias in the achieved sample.
- Methods of conducting research with organisations may not be well-established in all participating countries (e.g. there may be no suitable sampling frame of organisations and establishments).
- Such a sample clearly excludes those not currently in employment and a majority of the self-employed.

37. The third possibility for matching employer employee data would be to survey individuals and then to contact the employers of a random selection of these persons. The enterprise characteristics of the contacted employers could then be surveyed. An advantage of this approach is that it does not compromise the data characteristics of the probability sampling of individuals. These probabilistic features might not be present among employees sampled on the basis of a first selection of firms. Furthermore, beginning with a sample of employees – then working back to their employees – does not exclude the self-employed. The self-employed would be excluded from an approach that begins with the employer (across OECD member countries, approximately 13 per cent of the non-agricultural labour force is self-employed).

38. One example of such an approach is the UK Employer Perspectives Survey, which is an employer survey matched to the 2001 Skills Survey. The main objective of this survey was to investigate the relationship between the characteristics of the establishments where people work and the skills needed

in jobs. While the questionnaires are generally designed to capture the factors underlying the employer's demands for skills, they also aimed to address: (a) the relationship between skill requirements and the specification of the product or service generated at the establishment; and (b) the roles of technology, of competition, and of several modern management practices in influencing skill requirements and skills gaps (Green, Mayhew, and Molloy, 2003). 3,254 employers were identified from 4,470 individuals covered in the 2001 Skills Survey, from which 1,114 productive interviews were achieved.

Question for IEG members: should PIAAC include a survey of employers ? If so, should employee information be gathered starting from a sample of employers, or should a survey of employees be the basis for a selection of employers ?

39. While the job requirement approach and employer surveys are likely to provide rich sources of information, all existing surveys of these sorts exclude the non-employed. However, including the non-employed would shed light on a number of critical policy issues, such as: how different competencies relate to employment probabilities at different stages of the life-cycle; how different competencies relate to the duration of unemployment; and how competencies erode during time spent out of the workforce. It is thus important to consider ways in which the non-employed can be assessed.

6.3 *Testing the generic skills of the adult population*

40. PIAAC would assess selected generic skills of the adult population. The remainder of this section refers to a range of competency domains that could be considered for assessment under PIAAC. These domains have been selected either because their critical role in determining individual, enterprise or societal outcomes is well established – such as for literacy and numeracy - , or because they concern fields in which evidence of their critical importance is emerging – such as in team-working and high-level communication -, or because related issues have been raised by members of the IEG – as in the case of competencies and social capital. The choice of the competencies to be measured would of course depend on the availability of appropriate measurement techniques – including techniques that can be efficiently administered through a household survey. It is clear from the discussions below that valid and reliable cross-country and cross-cultural measurement instruments are not yet available for some key assessment domains. Table 1 presents an overview of the available assessment methods for various of the competency domains that could be measured under PIAAC (why these competency domains may be important for individuals and society as a whole is considered in the text that follows Table 1). IEG members are invited to consider whether there are any competency domains that should be addressed by PIAAC but which are missing from Table 1.

Table 1. Overview of Available Assessment Methods for Possible Competency Domains

Domains	Direct Assessment (Tests)	Indirect Assessment (Job Requirement)
Literacy / Numeracy	Exist for prose, document and quantitative dimensions of literacy - IALS, ALL, PISA	
Health literacy	Health Assessment Literacy Scale - ETS	
ICT	A framework and feasibility study successfully completed under PISA	Already used - UK Skills Survey
Team-working	Not currently available	Already used - UK Skills Survey
Communication	Not known	Already used - UK Skills Survey
Problem solving	Exist in national settings for different dimensions of problem solving. PISA and ALLS have developed tests for international assessment.	Already used - UK Skills Survey
Non-cognitive skills	Not available for international assessment	Not available for international assessment
Social capital	Not available	Not available

Question for IEG members: which of the competencies described above do members wish to assess in a first PIAAC cycle and why?

Question for IEG members: which of the competencies to be assessed should be assessed using the job requirement approach and which using other forms of assessment?

Question for IEG members: for which competencies – whether mentioned in this section or not – would there be interest in a programme of research aimed at the future development of measurement instruments ?

41. It is relevant to note that the DeSeCo project (Definition and Selection of Key Competencies) - which was sponsored by the governments of Switzerland and the United States in the context of the OECD International Indicators of Education Systems (INES) Project - represents a considerable intellectual investment in defining competencies that should be considered essential for individual and societal success (Rychen and Salganik, 2003 and Swiss Federal Statistical Office, 2003). DeSeCo has highlighted three broad categories of competence that together provide an overarching theoretical and conceptual framework for defining key competencies and affording high-level (pre-operational) guidance for their assessment. The three categories are: *Interacting in socially heterogeneous groups, acting autonomously and using tools interactively*. Various of the competency domains outlined below can be subsumed under at least one of the DeSeCo categories.¹⁵ For instance, different dimensions of literacy, numeracy and ICT competencies

¹⁵ Indeed, there are elements of overlap in the competencies that comprise each of the DeSeCo categories, such that a single competency domain may fall under more than one DeSeCo category.

relate to individuals' ability to use tools interactively. Literacy, numeracy and ICT competencies also relate to individuals' ability to act autonomously.

6.3.1 *Literacy/numeracy*

Motivation

42. Competencies in literacy have fundamental consequences for individuals, society and macro-economies. Literacy is a foundation competency. It facilitates the acquisition of information and additional competencies – including through self-directed learning and learning through electronic media – and is thus vital to social inclusion and life-long learning. Higher levels of literacy are associated with higher employment probabilities, reduced unemployment probabilities, a greater likelihood of being employed in a white-collar high-skilled occupation and, across many countries, higher earnings (OECD, 2000). For instance, research based on Canadian data shows that between a third and a half of the effect of schooling on earnings results from the increases in literacy brought about by schooling. There is also some evidence that the importance of literacy as a determinant of earnings has increased over the past thirty years (Green, 2001). Other non-market benefits of greater literacy may include superior health outcomes (see below for specific observations on health literacy) and higher levels of public and civic participation (including greater political participation among women as a country's average literacy skills rise).

43. At a macro-economic level, recent research using data from the IALS survey has identified critical linkages between literacy and growth (Coulombe *et al*, 2004). For instance, differences in average levels of literacy and numeracy were seen to explain 55% of differences in GDP per capita over the period 1960-1995. If the literacy/numeracy-growth relationships observed during this period were to remain stable into the future, then a 1% increase in average literacy and numeracy would yield a 1.5% permanent increase in GDP per capita. Consequently, even small increases in the middle of the literacy competencies distribution, where most workers are, would yield sizeable growth effects. Significant economic gains could be had from raising literacy among those with the lowest literacy competencies.

Assessing literacy/numeracy

44. IALS has, for the first time, collected reliable and internationally comparable data on the levels and distributions of broadly defined literacy competencies in the adult population. Combining a large-scale household-based assessment with educational testing techniques, across different languages and cultures, IALS represents an innovative development in international literacy assessment that could serve as a point of departure for an assessment of literacy under PIAAC. Indeed, for countries that participate in both IALS and PIAAC, adopting a similar framework to IALS for measuring literacy would allow examination over time of a range of policy issues associated with literacy.

6.3.2 *Health literacy*

Motivation

45. Health literacy refers to the capacities of individuals to “obtain, process, and understand basic health information and services needed to make appropriate health decisions.” (Rudd *et al*, 1999). Large numbers of adults possess a level of literacy below the reading requirements of health-related documentation, especially among at-risk population subgroups. When health literacy is inadequate, access

to care can be curtailed¹⁶ - including screening and early detection¹⁷ - and the efficacy of treatment impaired. Lower functional health literacy may also be associated with higher overall costs in health care. Furthermore, deficient literacy skills give rise to ethical considerations in the context of procedures that require informed consent from patients. Indeed, the full impact of inadequate health literacy has not yet been measured (Rudd *et al*, 2004). As well as being directly relevant to social equity, these issues are of growing importance in the context of rapid population aging in OECD economies. Demand for health literacy also undergoes change, being affected by developments in health and medicine (and, for individuals, by changes in individual or family health status, the birth of a child, etc.).

Assessing health literacy

46. Tests for health literacy have been developed. These include, in the United States, the Rapid Estimate of Adult Literacy in Medicine (REALM) and the Test of Functional Health Literacy in Adults (TOFHLA). REALM takes less than five minutes to complete and score, and involves participants reading from an ordered list of 125 common medical terms. TOFHLA tests comprehension of actual medical materials, such as patients' rights and responsibilities as stipulated in a Medicaid application form. These tests do not determine the cause or type of reading difficulty. However, they are useful in detecting patients for whom standard care approaches and materials may be ineffective (Rudd *et al*, 1999). Nor do these tests of reading skills constitute an encompassing assessment of health literacy, which requires competencies in prose and document literacy, oral exchanges and quantitative skills. In this connection, a recent study involved the creation of the Health Activities Literacy Scale (HALS) (Rudd *et al*, 2004). Based on a coding of data from pre-existing literacy surveys - the US National Adult Literacy Survey (NALS) and IALS - HALS measures competencies in health literacy relevant to a broad range of health activities that adults undertake in daily life.¹⁸ Literacy data obtained by PIAAC - if gathered using a framework similar to that used in IALS - could be used to create an index of health literacy similar to HALS. However, a more purposive sampling of health tasks/materials could be beneficial in order to capture issues around health promotion, protection, prevention, care and navigation in more equal proportions.¹⁹

6.3.3 *Information and Communication Technologies Skills*

Motivation

¹⁶ For instance, Rudd *et al* (1999) cite research on 958 English-speaking patients presenting for non-urgent care at an Atlanta walk-in clinic. Almost half of those studied were unable or limited in their ability to understand directions for medication or hospital documents.

¹⁷ For example, there is evidence that less literate men tend to present for prostate cancer at a more advanced stage of the disease. Similarly, lower reading ability in women is associated with lower utilisation of mammography.

¹⁸ ETS has subsequently developed an assessment tool: HALS focused on health literacy items. This will go on line at ETS on Sept 8 2004 [in English] and in French during the following months.

¹⁹ Thanks are expressed to Dr.Rima Rudd of Harvard University for this observation. Dr. Rudd has also brought to the attention of the OECD Secretariat the importance of oral language skills in the health context. Adults receive public health, medical, dental and mental health information over the airways and in discussion with professionals. The oral exchange is of critical importance and is related to literacy competencies.

47. ICT competencies are crucial to knowledge-based economies. ICT is strongly linked to macroeconomic growth, with ICT usage accounting for as much as 0.4% points difference in GDP per capita growth rates between the US, Germany, France, and Italy for 1995-2002 (EIU, 2004). There is evidence of high labour market returns to ICT competencies (Dolton and Makepeace, 2004). Knowledge workers (including computer workers) comprised the fastest-growing occupational category in the United States and the European Union between 1992 and 1999 (OECD, 2001: 2002).

48. ICT is increasingly considered a generic competence that all adults and workers need to possess. Rather than being confined to a relatively small sector of highly skilled information technology experts, the direct impact of computers has now spread across a diverse range of jobs. For instance, about 70% of employees in the UK use some type of automated or computerised equipment, and around 40% of employees judge computerised equipment as essential to their work (Felstead *et al*, 2002).

49. As ICT is a rapidly evolving field, forecasting which specific ICT competencies are likely to be in high demand in the future is problematic. Nevertheless, there is evidence that ICT-using firms in the US, Canada, Ireland and the UK tend to seek the following types of competence: technical ICT competencies (quantitative analysis, data modelling, digital media, technical writing, etc.) and business management competencies (marketing, strategy development and business writing) (OECD, 2002).

Assessing ICT competencies

50. Three methods relevant to assessing ICT competencies are considered here: formal qualifications, direct tests and the job-requirement approach:

51. Indirect information on ICT competencies can be gained by assessing formal qualifications in ICT. For example, as regards formal tertiary-level qualifications, the Taulbee Survey covers US and Canadian PhD-granting departments in computer science and computer engineering (CRA, 2000 and 2001). Private-sector certification might also be considered. For example, in the United States there has been an increase in the technical credentials granted by companies, business associations and commercial IT bodies (OECD, 2002). Around one seventh of ICT positions in the US now require commercial certification (Adelman, 2000). However, it is not known what such qualifications signify in terms of competencies in an international comparative context.

52. Direct tests of ICT competencies can also be administered. However, while ICT competencies have been tested directly, success has been mixed (Green, 2004). The UK Skills for Life survey 2003 also assesses individuals' use, purpose of use and intensity of use of various ICT competencies (the UK survey also asked these questions of non-workers).²⁰ It was intended that the ALL survey provide a test of ICT literacy, but an instrument was not finally developed. Instead, a small number of ICT-related questions were posed in the background questionnaire²¹. As a part of the PISA survey, in early 2003, an expert panel was established to address the issue of introducing a measure of ICT literacy.

²⁰ Many of the respondents indicated a relatively high level of awareness of ICT applications and terminology. However, good awareness did not always accompany good practical skills. Fifteen per cent of respondents had never used a computer and slightly fewer than half (47 per cent) achieved Level 1 or above in the practical assessment. Those who performed better were likely to use a computer most days, whether at home or work, or to use a computer for a variety of applications and were confident in their skills. Most accurately predicted their ICT skills, even though many adults over-estimated their levels of literacy and numeracy.

²¹ Appendix A provides further details on ALL.

The Panel presented a framework for ICT literacy skills in April 2003 (Lennon *et al*, 2003). After subsequent feasibility work, the Panel concluded that a computer-based assessment of ICT skills is both feasible and would allow the generation of new knowledge that could not be obtained from other types of assessment.

53. Using the-job requirement measurement approach provides an alternative way to assess ICT competencies. The 2001 UK Skills Survey provides information on job requirements in ICT skills. This survey extracts information by occupation and industry, covering the type, intensity, complexity and importance of computing skills and Internet use in the workplace. To gauge the existence and degree of skill mismatches, the survey also asks respondents to assess: (a) whether more computing skills would improve their job performance; and (b) if their computing skills could be better used in another job (Felstead *et al*, 2002).

6.3.4 *Team-working competencies*

Motivation

54. Effective teams are central to the productivity, responsiveness and output quality of public, private and non-governmental organisations. Indeed, teams are fast becoming a central feature of most organisations. To take just one example from the business world, Gordon (1992)²² found that 82 per cent of companies in the United States with 100 or more employees use some form of team. And a variety of studies in enterprise demographics show that entrepreneurial teams are associated with more rapid company growth and greater enterprise longevity. As new technologies and organisational forms augment the flow of information, the importance of teamwork is likely to grow. Indeed, Green *et al* (2001) provide evidence that competencies in team-working are of increasing importance in the United Kingdom.

Assessing team-working competencies

55. Feasibility work was initiated at the ETS, in the context of the ALL survey, to develop comprehensive instruments to assess teamwork (Baker *et al* [1999]). However, the results of this endeavour have so far proved unsatisfactory, principally because of great variation in teamwork attitudes and practices within and across countries (Murray, 2003). In the United Kingdom, Green *et al* (2001) have used the job-requirement approach to assess a small number of social and team-working competencies. Research would be needed to assess whether the job-requirement technique could be used to measure this competency domain internationally, or whether any assessment method would likely incur the same difficulties as faced in the development and testing of measures of teamwork under the ALL survey.

6.3.5 *Communication competencies*

Motivation

56. There is evidence of significant labour market returns to competencies in high-level communication, such as the ability to write reports, persuade and influence others, and make presentations and speeches (COM/DELSA/EDU(2004)6 and Green *et al* (2001).

Assessing communication competencies

²² Cited in Baker *et al* (1999).

57. Some components of communication competencies are likely to be captured in measures of literacy. Correlations between literacy and other facets of communication competencies would need to be further investigated. However, a wide range of communication competencies could be measured using the job-requirement approach (see Green *et al* [2001]).

6.3.6 *Problem solving/analytical reasoning competencies*

Motivation

58. Future-oriented studies of work-force competency requirements identify problem-solving as a core competence (2003a). Indeed, problem-solving / analytical reasoning is a competence that may be integral to all other domains of competence. That is, competencies in problem-solving may underpin the use of diverse tools (such as language, number, team-based collaboration and new technologies) in seeking solutions to an array of problem types.

Assessing problem-solving/analytical reasoning

59. The assessment of problem-solving/analytical reasoning has generated a vast literature. However, there appears to be no consensus on the definition of problem-solving. Tests for problem-solving relevant to specific problem types have been developed in national contexts. Until recently, frameworks for assessing problem-solving as a cross-disciplinary competency have been lacking (OECD 2003a). However, both ALL and PISA have developed frameworks and instruments for the international assessment of problem-solving. While broad, the problem-solving framework and measurement instrument developed for PISA 2003 does not cover all dimensions of problem-solving. For instance, it excludes interpersonal and group problem solving, which are considered important by employers. By contrast, the job requirement approach has been used to assess problem solving in the workplace. Given that the results from tests of analytical reasoning and problem-solving may be highly correlated with numeracy and literacy (data collected under PISA 2003 and the ALL survey will help to quantify this relationship), assessing problem-solving using a job-requirement approach could capture new information in a wider set of real-world problem situations (see Green *et al*, 2001).

6.3.7 *Non-cognitive competencies*

Motivation

60. Analysts have recently given increased attention to the economic role of non-cognitive competencies such as creativity, motivation, reliability, self-discipline, tolerance and willingness to work in a team. Surveys indicate the essential importance that employers accord non-cognitive competencies. For instance, in the United States, a 1998 national survey of employers found that the two most important characteristics sought in new non-supervisory recruits were good “attitude” and “communications skills” (Bowles and Gintis, 2000). Similarly, Green *et al*. (1998) describe the causes of skills shortages identified in a survey of British employers. As a cause of skills shortages, poor “attitude, motivation or personality” among workers was cited by 62 per cent of employers. Only 43 per cent of employers referred to deficiencies in technical skills. Conversely, data from the United States suggest that cognitive skills are a weak predictor of wages (Cawley *et al*, 1996). There is also evidence that non-cognitive competencies may account for anywhere from 55 to 80 per cent of the returns to schooling (Green, 2001). Green (2001)

likewise suggests that non-cognitive skills are rewarded in their own right, independently of their role in raising the productivity of cognitive skills.²³

61. Heckman and Carneiro (2003) demonstrate the importance of cognitive and non-cognitive skills formed early in life in accounting for key dimensions of socioeconomic success. They also describe the generally positive evidence for the effectiveness of adolescent mentoring policies in the United States. The success of these policies is largely a result of improvements in social competencies and motivation. If non-cognitive competencies could be assessed, PIAAC might shed light on whether different features of schooling and training are more or less effective in producing such competencies. The degree to which non-cognitive competencies in adults are affected by work experience or by policy is an open question (however, in the United Kingdom, Green *et al* [2001] show that a significant share of the variance in team-working and social skills is associated with firms' organizational characteristics and indicators of work-based learning).

Assessing non-cognitive competencies

62. Owing to the fact that a diverse set of traits is included in the category "non-cognitive competencies", measurement is problematic. Tests of non-cognitive competencies have been used by companies to screen employees, but have not been used to assess the effectiveness of schools. And it is unlikely that any single measure of non-cognitive skills could be found (Heckman and Carneiro, 2003). A large number of studies have linked measures of such personality traits as self-efficacy, locus of control and self-esteem to labour market outcomes. However, the relevance of such research to policy appears unclear. Nevertheless, tests relevant to a small number of non-cognitive skills have been used and/or are being developed. For instance, researchers at the US Education Testing Service, in the Natural Language Processing group, are working to develop methods for detecting creative writing in automated scoring programmes. And attitudes to teamwork - a behaviour resulting from a number of individual non-cognitive skills and attitudes (as well as organisational settings) - are open to measurement (Baker *et al* [1999]).

6.4 Other information to be collected on respondents

63. Rich contextual information is vital to the analytical power of PIAAC. Thus, careful consideration needs to be given to the types of background information sought from respondents. The appropriateness of the background questions will depend on participating countries' decisions regarding the key policy goals of PIAAC and the competency domains and age foci to cover. As an indicative outline, consideration might be given to questions that elicit information on different dimensions of respondents':

- Labour market, socio-economic and demographic characteristics (including parental, home and linguistic background);
- Attainment in initial education, adult learning experience and when and where this occurred (including formal and informal training; work-based and non-work-based training);
- General and work-related literacy and numeracy practices;

²³ However, certain non-cognitive skills will only be utilised if organisational structures and work practices permit. In practice, many employers pay lip-service to concepts of teamwork, creativity and individual responsibility.

- Workplace characteristics (such as whether workers have opportunities for expressing views to management, whether there are performance appraisal systems in place, quality circles, etc);
- When other suitable sources of information on employers are unavailable, characteristics of employing firms (e.g. firm size and industry);
- Personal use of ICT and a range of related or other technologies;
- Practices related to social capital, civic participation and physical well-being.

Question for IEG members: which of the above and other possible items of contextual information are considered of greatest interest?

6.5 Social capital and citizenship

64. Various participants at the first meeting of the PIAAC IEG expressed interest in assessing adult competencies in relation to social capital and citizenship (COM/ELSA/EDU(2004)7). In this connection, a first issue to consider is: What is social capital? A definition proposed in a frequently-cited OECD report on the subject is: “*networks together with shared norms, values and understandings that facilitate co-operation within or among groups*” (OECD, 2001). However, the term “social capital” encompasses a set of interrelated but diverse concepts, some of which posit different causal mechanisms to economic outcomes. Indeed, while social capital is increasingly viewed in terms of networks and civic norms, the definition and *modus operandi* of social capital does not command a consensus among theorists.²⁴ Nevertheless, in recent years social capital has attracted considerable media and policy attention. The potential benefits of increased social capital have been documented in a now copious literature. To take just one example, by increasing the availability of information and lowering some economic transaction costs it is clear that greater connectedness arising from various forms of networking could have positive individual, economic and social benefits.

Assessing social capital

65. Difficulties of empirical measurement are a key concern in academic and policy debates on social capital. These difficulties reflect both the limited tangibility of some features of social capital and the lack of an accepted definition of the concept itself. For instance, trust is the norm most frequently cited in connection with social capital. Surveys exist of inter-personal trust, such as the *World Values Survey*.²⁵ But these address only a single facet of social capital (as well as being open to conceptual and methodological critique). Furthermore, measures of network activity vary in their significance from one realm of economic or social life to another. For example, business-oriented networks maintained by entrepreneurs differ greatly in both function and form. They differ still further from the social networks maintained by groups such as the long-term unemployed. Some aspects of social capital can be quantified – such as the extent of

²⁴ Ritzen and Woolcock (2000) note that even the term “capital” can be confusing when referring to social issues. For instance, important characteristics of physical capital do not apply to social capital, such as divisibility and the possibility of establishing ownership.

²⁵ See <http://wvs.isr.umich.edu/index.html>

volunteering, and the degree of participation in civic organizations. However, at present, empirical measurement is a contested subject.²⁶

66. Critically, it is also unclear which individual competencies contribute to different strands of social capital and in what ways. For instance, whether competencies are connected with values and norms in any systematic way is unknown.

67. For the purposes of PIAAC, developmental work for an assessment that incorporates social capital could be resource-intensive, with grounds for doubt over whether useful survey instruments would result. However, even without new instruments to assess competencies in the context of social capital, PIAAC would allow the analysis of relationships between the stocks and distributions of adult competencies and data on important concepts that are often referred to in the context of social capital. These could include various dimensions of citizenship and social cohesion (the latter understood as societal inclusion and political responsiveness [Ritzen and Woolcock, 2000]). For instance, data collected as part of the IALS survey indicate a relationship between literacy and political participation (OECD, 2000). This relationship could be explored in more detail. PIAAC could also relate labour market success - the principal determinant of social inclusion - to a wide range of issues in the development of competencies. Furthermore, as part of the background information on PIAAC interviewees, countries might consider whether to include questions relevant to selected aspects of citizenship, social cohesion (or specific and tangible features of social capital). These might include such behaviors as voting frequency, the extent to which television is a primary information source, time allocated to voluntary or charitable activities, active membership of clubs and associations, etc.

6.6 *The Number of Skills Domains to be Assessed*

68. There would be advantages to an assessment that covers a wide range of competencies. For instance, testing over a larger number of domains would allow the examination of relationships between competencies and labour market and education policy for which evidence is not yet conclusive. New and possibly unexpected findings could also result. Furthermore, it is relevant to assess the extent to which individuals in different countries and labour markets possess multiple competencies (Green, 2004). However, measuring a larger number of competencies could give rise to operational and financial constraints. Murray (2003) notes that the IALS experience indicates that respondents in a household survey will generally not undertake an interview lasting more than 90 minutes. In this connection, the overall cost implications of increasing the range of competencies measured will depend on the measurement techniques used. Using the job-requirement approach, the range of competencies assessed could be expanded without a major increase in cost, but doing so may make the duration of an interview excessive. It is possible to ask respondents to do long interviews, but a tired or resentful person does not make a good respondent. One solution here is to increase the overall sample size and split the data collection tasks in some way, which can help to increase the response rate. But this option is more costly than taking a smaller sample in which people are asked 50% more questions, and is also more difficult to analyze. Paying a financial incentive for

²⁶ Even if valid, reliable and accurate measures of social capital were available they would not provide unambiguous linkages to (desirable) policy outcomes. For instance, it has been observed that certain forms of social capital could have pernicious consequences, for instance by locking-in or facilitating dysfunctional norms, or by stifling challenges to outmoded forms of economic organisation. Indeed, an early international quantitative study of social capital (Knack and Keefer, 1997) found that negative economic effects were associated with higher levels of associative activity (by way of explanation, the authors hypothesized that in societies marked by important social, economic or other divisions, greater intra-group networking may serve to entrench inter-group conflict).

the longer interview may be a worthwhile compromise.²⁷ Also, Murray [2003] notes that measuring competencies in each domain, using test booklets for each domain, as well as an inter-domain covariance matrix, significantly increases the required sample size, with a concomitant increase in costs²⁸.

6.7 *Preparing Complementary Qualitative Studies*

69. Quantitative measures of key competencies could be supplemented with a focused and co-ordinated programme of qualitative studies. Qualitative studies can be particularly useful in revealing information about the processes that underlie relationships in quantitative data. In this connection, there are a variety of policy themes on which qualitative studies could improve understanding. Green (2004) notes that policy insight could be gleaned through qualitative studies of how recent graduates utilise competencies in the workplace. Such studies are rare in an international comparative context. They would also be highly relevant to policymakers concerned with issues of investment in tertiary-level education. If undertaken, a series of targeted studies should be performed in a number of countries using comparable methods and covering agreed sectors and enterprise types. Feasibility work would be required to consider such issues as the preferred number of participating countries, the choice of policy issues to address, and study methods.

Question for IEG members: do members consider that complementary qualitative studies could add value to a first assessment cycle? If “yes”, which subject areas might be of greatest interest ?

7. **Links to subsequent assessment cycles**

70. The text so far has concentrated on presenting the strategy choices that would need to be made for a first PIAAC assessment cycle. It follows that making these choices – over age foci, skills domains, complementary studies, sampling strategy, etc. – will begin to set a framework for new choices regarding subsequent PIAAC cycles. Even at this early stage, it is important that members of the IEG begin to specify goals for PIAAC beyond the first cycle, and to decide how to sequence policy priorities over time. IEG members are therefore invited to reflect on how preferred choices for a first assessment cycle would relate to the expected objectives of subsequent cycles.

71. One obvious linkage between PIAAC cycles would result from choices of survey method, particularly if a goal was to track the same individuals over time through a longitudinal survey. In this

²⁷ It is worth asking if all topic areas require the same sample size. If some of the competencies are found among a small minority of the population/workforce, then a large sample size is required to generate enough material for analysis of this competency. That may mean that the remaining sample members are more numerous than required for many of the other question areas. However, there may be different areas of competency which are more relevant for certain sub-groups. For example, if someone has a degree, there is little point in getting this person to sit an exercise designed to evaluate basic literacy. The key to respondent co-operation is for them to feel the questions asked are relevant to them. This could mean that an overall 90-minutes of question material is compatible with an average interview length of 60 minutes.

²⁸ Indeed, the PISA experience indicates that matrix sampling may address the problem of scope. Up to 7 hours of content was achieved without exceeding 2 hours of individual testing time.

connection, consideration is given here to the practical and conceptual merits and demerits of a longitudinal survey.

7.1 *Merits and demerits of a longitudinal survey*

72. The issue discussed here is the implications for survey method and cost of a longitudinal survey (referred to here as a 'panel survey'), as compared with a series of cross-sectional studies. The premise of conducting data collection at a series of points in time is that one is interested in change. Starting in a known position, how far has the position changed over a period of time? A survey estimate includes an element of error. Although statistics can help to understand the probable magnitude of the sampling error, this can only be estimated. If repeated cross sectional surveys are undertaken, what appears to be a change could in fact be a consequence of the estimates varying in a way that appears to indicate change.

73. A large sample size contributes to statistics that show a smaller margin of sampling error. However, in the context of adult competencies, the required sample size is likely to depend mainly on the requirement for subgroup analysis rather than gaining precision for estimates for a country's economically active population. This is because the overall estimate is likely to be of less interest than making comparisons by region, occupational group or age group, both within national entities and between them. It is also arguable that adult competencies are difficult to measure exactly, which means that a degree of non-sampling error cannot be avoided. As a result, there could be uncertainty about the magnitude of change in competencies, even if a lot of money has been spent on obtaining a very large sample.

74. How hard it is to measure a rate of change depends partly on how large it is expected to be and the period of time observed. One advantage of a panel study over a cross-sectional study is that (in principle) the error is reduced in estimates that relate to the sample design. For a given sample size, this allows us to detect small changes with greater confidence than we could with two cross-sectional studies of the same size (even if both the cross-sectional studies had the same sample design and questionnaires). It may be worth noting that many sample members may have changed jobs over the space of three or four years, and this sort of change may be of interest in itself; the design of the Labour Force Surveys in Europe only captures details of both current and prior jobs for short-term job changes. However, some of these changes of employer may not have involved a change of occupation, so the amount of change in competencies may be smaller in magnitude. If changes in jobs are an important aspect of the research, then retrospective collection of work histories may be useful.

75. It may help to clarify these issues by looking at a specific example. Suppose in 2006 a survey has been conducted of 4,000 people aged 20 to 60 and in work at the time of interview. The issue is whether in 2009 it is preferable for to re-interview the same people or to obtain a fresh sample in order to assess the changes over this period.

76. To prefer the panel sample to estimate changes in adult competencies, it has to be believed that the measurement process is sufficiently accurate to enable the identification of aggregate changes for each

sub-group of interest.²⁹ This also has to hold true when the sample size is reduced to the people interviewed successfully on both occasions.³⁰

77. With a panel design, one can expect to find a majority of the sample members still living at the address where they were interviewed originally. These people may be relatively easy to interview, but their lives may have changed little over the period of three years. A significant part of the change to be estimated is likely to be found among the people who have moved (and the rate of movement may be greater if the original sample was of the younger age cohort). Efforts to locate people who have moved by 2009 may well involve greater unit costs than the original work of making contact with them in 2006. Many may have moved to localities that were not among the sample points used in 2006, and this means that the costs of tracing and interviewing may be relatively high. Shortening the time interval between surveys is the best way to alleviate attrition due to difficulty of tracing panel members; but then the extent of change may be too small to estimate with confidence. If one allows a long enough period for changes to occur, the high rate of attrition may mean that one can estimate the amount of change for only a fraction of the people whose behaviour is of greatest interest.³¹

78. The alternative is to conduct a second survey with a fresh cross-sectional sample and infer changes by comparing the findings. One can minimise the effect of sampling error by maintaining continuity in the sample design – for example, selecting a new set of addresses within the same Primary Sampling Units (PSUs, postcode sectors in Britain). One can also avoid a likely source of spurious change by using the same organisation to conduct data collection, since it will use consistent quality control methods and the same software for interviewing. In many cases, the same interviewers may be available on the second occasion.

79. Another way to mitigate the effect of a fresh cross-sectional sample on the measurement of change is to ‘match’ people in the second sample to their counterparts in the first sample.³² At one level, this can involve weighting the profile of the second sample to be an exact replicate of the first. One might take this further, and reduce the sample to a set of people for whom exact counterparts can be found at both stages. Clearly, one would not treat these pairs as showing the change in competencies at an individual level; but one is probably constrained, in any case, to do the comparison of competencies over time for sub-groups of the sample. In effect, with this approach, we accept that the original estimate may have been biased to some extent, but we minimise the chance of the bias differing from one stage of the study to the next. At the same time, a fresh sample is more likely to be representative of the cross-sectional situation at the time than a follow-up of the original sample is likely to be (and one would of course use all the cases for this sort of analysis, rather than the matched sample).

²⁹ It is interesting to consider whether the method is viewed as being sufficiently robust to look at any individual and estimate the change in their competencies. I am assuming that this is not the sort of analysis being considered, but one needs to have a view on this to assess whether aggregate estimates are likely to have a large element of random variation.

³⁰ Inevitably, even if we manage to trace most of the people who have changed where they live, there will be other losses, due for instance to some people becoming economically inactive (e.g. reaching retirement age).

³¹ Birth cohorts spend large amounts to minimise attrition, but still have to accept a high rate of loss, especially in the earlier stages of a study.

³² There is a large literature on matching methods, which are not considered in more detail in this paper. One might wish to match cohorts who experienced educational systems and the economic cycle in the same way, so one could choose to match a person aged 27 in the first survey to one aged 30-31 in the second.

80. Another merit in selecting a separate sample is that potentially the two samples can be combined to allow examination of sub-groups that were too small for reliable analysis on the basis of a single survey. This would not be possible with a panel.

81. The limitations of a panel study would be reduced if one had a reliable and low-cost method of tracing members of the panel after a period of, say, five years. Countries which have national population registers (or access to something equivalent, such as social security numbers with addresses) might consider a panel to be the obvious choice of survey design. One can envisage a situation in which some countries preferred to have a longitudinal sample and others a repeat cross-sectional sample. These are two different types of sample, but they are not necessarily incompatible.

82. In other situations, a panel survey could be at least as expensive as a new cross-sectional study, largely due to the amount of effort required to trace movers. This is likely to be the case, even when effective methods have been used to keep in contact with panel members in the period between the surveys.³³

³³ Following the 2001 UK Skills Survey, as a measure to enable a follow-up study to take place later, sample members were re-contacted around 15-18 months after the original interview, receiving a short self-completion questionnaire with a short account of the main findings of the original study. Without resources for a very intensive follow-up (e.g. being able to pursue non-respondents through 'stable addresses' recorded at the initial interview), a response was obtained from 80 per cent of the original sample members (around 90 per cent of whom had given permission to be re-contacted). Developments in mobile phones and Internet are changing the mechanisms available for keeping in contact.

PART C. A FIRST CONSIDERATION OF RESOURCE REQUIREMENTS

8. Approximate cost implications of different strategy options

83. Four potential age-related assessment options are considered here:

- A focus on youth (18 to 35 years of age³⁴),
- A focus on older adults (45 to 65),
- A focus on the entire age spectrum (18 to 65 or 20 to 60),
- A focus on the entire age spectrum with over-sampling of a selected age group.

84. The resource implications associated with each of these options is the main focus of the discussion in this part of the paper (however, very approximate information is also provided on the resource requirements that would be entailed should PIAAC seek to gather data on employers).

85. For the purpose of this discussion, it is envisaged that an interview would be conducted by a survey research interviewer at the respondent's home. The interviewer would have been given a day's training³⁵ in how to handle the study before starting work. It is also assumed that all assessments will be handled by the interviewer, without requiring a high level of skill in the assessment task. As a result, the amount of time needed to conduct the interviews is assumed to be fixed across the four strategy options. Hence, the variable element is the amount of time needed to make contact with the sample member, which is likely to vary according to the age range of interest.³⁶ The estimates of resource requirements considered here are based on data from the United Kingdom. How such estimates are likely to vary across countries is considered later, in section 9.1.

86. A classification of the different types of major costs that would be incurred under PIAAC is set out in Appendix B. These cost categories include: survey administration; sampling; questionnaire programming; briefing of interviewers; conducting data collection; pilot survey work; data preparation; data weighting; preparation of technical reports; and project management.

87. Nationally-representative surveys of this type in the United Kingdom are based on probability sampling methods, using a database of postal delivery points recorded by the Post Office. In the past, the Electoral Register (all persons aged 18 and over who had registered to vote) was commonly used for sampling. Using these sampling frames, one knows little or nothing about the eligibility of the residents at an address until the interviewer manages to make contact with them. Making contact often requires a

³⁴ The exact age range has not been specified – this is suggested as the basis for estimating costs.

³⁵ The interviewer would previously have been trained over 2-3 days in basic skills of conducting structured interviews using Computer-Assisted Personal Interviewing (CAPI).

³⁶ The most important exception to this occurs in countries that have registers of individuals, in which case a pre-selected, named sample can be used.

number of visits and for this reason the sample is clustered into interviewer assignments, each of which is usually expected to yield around 20 productive interviews.

88. In the United Kingdom in 2001³⁷, it was found that 54 per cent of residential addresses contained someone aged 20 to 60 who was currently in paid work. Approximately 20 per cent of addresses in the United Kingdom contain only people aged 66 and over. As these addresses cannot be identified in advance, the interviewer has to make contact and identify whether anyone resides at the address who meets the age criteria for the study. However, not all of the addresses which contain someone aged under 65 are going to be eligible. In some cases, the other residents are too young. In others, they may not be in paid work (employed or self-employed) or have not worked for a period in the past. They may be of the correct age but are long-term unemployed or have a long-term health condition or care for young children or adult dependents.

89. As well as the age range to be covered, some decisions are required about who meets the criteria of interest. Ultimately, this is a matter of how much a wider definition serves the aims of the research. This section aims to indicate the implications of different ways of defining the sample in terms of the costs of making contact with sample members and the costs of conducting an interview. As previously discussed, given that work-related competencies are of particular interest, one might decide, as in the 2001 Skills Survey, that only people currently in paid work are of sufficient interest to be interviewed.

90. There is a further important source of costs in conducting surveys. In the United Kingdom in 2001, the response rate was 64 per cent. Thus, three eligible people had to be found for each two people interviewed. The number of addresses to be contacted by interviewers was approximately three times larger than the number of interviews achieved. Response rates vary between countries, so the expected level of response is likely to be an important part of the translation of these U.K. examples into estimates for other countries.

91. The choice between the various strategy options involves similar issues. The narrower the age range of interest, the greater is the cost of making contact relative to the cost of conducting interviews. The figures quoted here are approximate and may well vary from country to country, but for the sake of illustration, it is assumed that:

- For the age range 18 to 65, it is necessary to contact around three addresses to find someone in the correct age range at two addresses,
- For the age range 20 to 60 and a further criterion of being in work, it is necessary to contact two addresses to find one eligible person,
- For the age range 18 to 35, it is necessary to contact around six addresses to find one person in the correct age range,
- For the age range 18 to 27, it is necessary to contact around ten addresses to find one eligible person.

92. At some point, a position is reached at which the costs of making contact are so great that one should look for another way of locating the sample members.³⁸ The technique known as ‘focused

³⁷ Felstead *et al*, 2002, Work Skills in Britain, DfES.

³⁸ This is not an issue in any country where sampling is based on population registers.

enumeration' was developed in the 1980s (originally to identify members of ethnic minorities in the United Kingdom). This relies on people's knowledge of other people living in neighbouring accommodation. The interviewer makes contact at one address, but asks about the occupants at the address itself and at one, two or three neighbouring addresses to each side.³⁹ This improves the ratio of interview time to contact time, but not as much as one would expect – the information obtained is often only about 70-80 per cent reliable. The reliability of this method is probably highest (80 per cent or better) when someone at the middle address is eligible for interview. In this case, the question about neighbouring addresses can be asked after the person has been interviewed and has had time to come to trust the interviewer. For instance, in the case of the strategy focus on the entire age spectrum with over-sampling of a particular age group, one would always approach the middle address, and would conduct an interview there if someone aged 18 to 65 was found. The three neighbouring addresses on either side would be 'screened' for the presence of 'young' people (say, aged 18 to 35), and one might expect, on average, around one interview per set of six addresses.

93. Contrasting cost estimates for data collection are shown in Appendix D. These are based on the sampling methods used in the United Kingdom, namely starting with a probability sample of addresses. These addresses are clustered in a large number of postcode sectors.⁴⁰ The cost estimates correspond to each of the four possible age-foci options discussed previously in Part B of this paper. A selection is made of sets of 5 adjacent addresses and one person resident at any one of the five addresses may be asked to say which of the other 4 addresses they believe to have no-one eligible. These addresses are then removed from further effort by the interviewer. Where the person spoken to was unsure, or if someone aged 18-35 was identified, then the interviewer conducts further contacts in the usual way.

94. The cost estimates do not include the costs of developing the questions to be used for the assessment of competencies. Considerations relevant to the estimation of these costs are set out in Appendix A. Nor do these cost estimates include resources needed for the analysis of collected data. Table 2 summarises the cost estimates.

Table 1. Approximate Survey Costs (in Euros) for Four Possible Age Foci

Approximate cost				
Achieved Sample Size	All adults aged 18-65	Adults aged 18-35	Adults aged 45-65	Adults aged 18-65 with 100% extra 18-35s
2 500	873 450	1 039 200	1 039 200	(sample size 3 500) 1 184 400
4 000	1 175 700	1 400 700	1 400 700	(sample size 7 000) 2 128 800
7 500	1 690 950	2 114 700	2 114 700	(sample size 10 500) 3 076 200

³⁹ In general, it seems best to use this method to identify addresses that are clearly ineligible, leaving the interviewer to make contact in person at any addresses that could be in-scope.

⁴⁰ A UK postcode sector has an average of around 2,500 units of accommodation.

95. Earlier discussion in this paper considered possible approaches to incorporating an employers' perspective in PIAAC. Precise costing of this option is not available at the time of preparation of this paper, and would require further study (the Secretariat is currently investigating this subject). However, **it is unlikely that the cost of interviewing a single senior manager at each establishment in a sample of 2,000 achieved interviews would be less than 1 million Euros.** If an interviewer was then going to attempt to conduct interviews with 20-25 employees, this would add around another three whole days on-site (8 interviews per day per interviewer, assuming each interview lasted 30-40 minutes). If each day cost 350 Euros (including coding the data, research coordination etc), this would add around 2.1 million euros. This could yield an achieved sample of 30,000 achieved interviews (assuming 15 productive interviews per site, on average), and hence a much lower cost per interview than a household survey. This is assuming that employers would be willing to allow so many of their employees to take part in the study. A more practical design could involve a smaller sample per workplace of maybe 8 interviews, based on some kind of random sample stratified by occupation. Based on the same costs assumptions, this would result in a cost of around 0.7 million Euros. In other words, **an extremely approximate cost estimate for a survey that encompasses senior managers in 2000 enterprises plus interviews with 8 employees in each workplace would be 1.7 million Euros.**⁴¹

96. The discussion in this section has focused on the density of eligible persons in the population. A separate aspect of the age group used for the research is the uniformity of the qualifications system across different age cohorts. Most activity leading to gaining qualifications occurs among young people. It follows that, on the whole, they are more reliable as informants about their qualifications than those aged 50 or over, who often gained their qualifications a long time in the past. Older respondents may be asked about the qualifications they obtained when they were young, but problems of memory may mean the data obtained are of poorer quality. It is possible that members of a country's sample may have been educated or trained under two or three different systems. The point here is simply that respondents should be asked about the system that applied at the time they gained their qualifications, and should not have to do the translation from one system to another.

8.1 *Factors likely to cause cross-country variation in costs*

97. High-quality survey research is costly, so perhaps the biggest source of variation in costs between different countries is varying conventions regarding quality standards. For example, suppose the expected response rate were around 50 per cent in a particular country. It is quite likely that the unit costs of data collection will be lower than would occur if additional effort was put into raising the response rate (say, to 70 per cent). However, this is not purely a matter of effort and cost. If the interviewer-dependent aspects of response have been dealt with effectively (in particular, the non-contact rate has been minimised), then most of the non-response is likely to consist of refusal by respondents. There may be cross-country differences in the willingness of members of the public to take part in such surveys, as well as differences in the extent to which it is considered acceptable to make a second attempt with sample members who refuse. Re-issuing refusals to more senior interviewers to follow-up is always a costly way in which to raise the response rate.⁴²

⁴¹ It is also worth bearing in mind that were PIAAC to include an employers' perspective, the range of possible funding institutions might be considerably enlarged. For instance, Ministries of industry, regional and enterprise development agencies and, possibly some private and non-governmental organisations, would have a natural interest in issues of competencies and enterprise-level innovation.

⁴² This is a different point from using re-issuing to a second interviewer when the original interviewer's efforts were below an acceptable standard. In this case, the re-issuing is a quality control measure.

98. While the sample design is intended to ensure that each country's sample includes all types of community, including those that are rural, the proportion of time an interviewer can spend productively is affected to a degree by the distances between addresses. Excessive clustering reduces the effective sample size (whereas appropriate stratification may have a beneficial effect), so there must always be a trade-off between the size of individual clusters and the efficiency of the way interviewers spend their time⁴³.

99. It is assumed that all countries have suitable research organisations. However, in some countries, the norm is to collect data by telephone and it is relatively unusual to interview face to face. As a result, it may be relatively costly to conduct interviewing in people's homes, simply because respondents do not expect this to happen and interviewers have little experience of working this way.

100. If interviewers in the organisation contracted to collect data are accustomed to market research (as opposed to social survey research), they may need additional training in sampling (e.g. respondent selection using a Kish Grid method) and additional supervision to keep trying when they have made a specified number of calls without yet making contact with the intended respondent.

101. One of the potential threats to data quality is the use of numerous different software systems⁴⁴ in different countries. However, it may be impractical to insist that all organisations use the same software for interviewing, which could imply that some have to license the software and develop support systems specifically for PIAAC. In this situation it is arguable that software development and sample administration should be carried out by an organisation located in a second country, which then has an involvement in the management of data collection and supervision of interviewers.⁴⁵ In effect, in this case, the fieldwork operation is being sub-contracted and other aspects of the project are being managed alongside the work in a different country. This arrangement could have beneficial effects on data quality. In this connection, Appendix B outlines a number of the implications and advantages of administering at least part of a survey using a computer.

102. It should be noted that market research now operates on a global scale. It is possible that there exist organisations with the capacity to administer an assessment of this sort in 20 countries or more. The chief problem with this could be the loss of autonomy among national research teams.

103. Appendix D describes a number of assumptions on which the costs of conducting a PIAAC survey in the United Kingdom have been estimated (based on costs in 2004-05). Underlying these estimates is a set of more fundamental assumptions that relate to institutional conditions and practices. Members of the IEG may be able to relate these assumptions to institutional practices in their respective countries and thus obtain an indication of whether the cost estimates given above (and in Appendix D) are likely to be relatively low or high for their countries. The key institutional conditions to consider are:

- *Does an organisation exist which has a suitable fieldforce?*
- *Does an existing organisation operate at acceptable quality standards?*

⁴³ Experience in the United Kingdom suggests that it may be more cost-effective to send an experienced interviewer to a remote community and cover the costs of travel and accommodation than to recruit someone from the area who will be able to work on few surveys.

⁴⁴ For example, few of the available systems are equally strong in collection of text, numerical and date format data.

⁴⁵ As an illustration of how this may work, the National Centre for Social Research provided the questionnaire for interviewing on a British Crime Survey to be conducted by a different research organisation in Northern Ireland, and also undertook specialised coding and editing of the data. However, both organisations in this case used the same CAPI software (Blaise).

- *Is it acceptable to ask a respondent to be 'tested'?*
- *What is the impact of literacy – including computer literacy - on the assessment of adult competencies?*
- *Is a high-quality sampling frame available?*
- *Does the research organisation have experience of high-quality coding to international classifications of occupations and industries?*

104. How these institutional conditions and practices relate to costs is considered in the following paragraphs:

Does an organisation exist which has a suitable fieldforce?

105. The underlying issue here is whether the assessment of adult competencies requires a particular type of skill. In other words, if 'ordinary' interviewers were used, would their lack of specific skills in assessment techniques run the risk of undermining the value of the research? This appears not to have been the case in the ALL survey. Suppose, however, that it was decided that interviewers would need 2-3 days of training to administer a series of tests. It is worth considering whether it might be more cost-efficient to recruit a fieldforce of people who already have the appropriate skills (e.g. occupational psychologists or teachers) and give them the training necessary to act as interviewers. In the end, the choice might be to use interviewers, whose rates of pay are generally below those of professionals. Another reason for doing this is that it takes most interviewers 2-3 assignments to acquire the skills needed to obtain a high response rate. A further issue is that the costs of recruitment and basic training should not be under-estimated. Thus, it is assumed that an existing fieldforce will be used in most countries, and the format of the assessments will need to be designed accordingly. However, some countries might decide to adopt the alternative approach, using as interviewers people with specialist skills in assessment.

Does an existing organisation operate at acceptable quality standards?

106. There are many aspects to survey quality. Even if the questionnaire design has been standardised across countries, there are other important ways in which quality can be compromised. These include, for example:

- Errors in the implementation of computer-assisted interviewing which can lead to loss of data (this risk can be minimised by thorough testing prior to fieldwork and rigorous quality control as data are returned.)
- Pseudo-random sampling methods, which are common in much market research. These lead to 'samples of convenience' which do not have a known relation to the population they are supposed to represent.
- Whether the interviewer has selected the right person (it will often be more profitable for them to 'cheat' on selection: that is, to select the person who happens to be available when they first make contact, even though a different person should have been selected if the sampling rules had been applied correctly).

Is it acceptable to ask a respondent to be 'tested'?

107. The answer might vary from one country to another. In Britain, it is increasingly common to give survey respondents a reward (often in the form of a voucher, which can be exchanged in many shops) for giving their time and attention to the study.⁴⁶ This is especially common with panel studies, so as to maximise the willingness to be re-contacted. In general, payment reduces the likelihood of people refusing to take part in the study. This could be important to help achieve a satisfactory level of participation by people with lower levels of competencies, who may be anxious about being 'tested'.

What is the impact of literacy – including computer literacy - on the assessment of adult competencies?

108. Survey response rates on social research studies in the United Kingdom vary from around 60 per cent to over 80 per cent, to a large extent related to the intrinsic interest of the topic. Health and crime surveys achieve high response rates. Social attitudes and political studies tend to have the lowest response rates. Skills, training and learning tend towards the lower end of this range (for instance, the response rate was 66 per cent for the 2001 UK Skills Survey).⁴⁷

109. Functional illiteracy affects an estimated 10 to 15 per cent of adults in the United Kingdom. It is highly likely that these people will refuse to participate in surveys, especially where the subject appears to relate to their (lack of) skills. The ethics of conducting surveys mean that efforts to persuade a reluctant person to participate have limits. A key skill on the part of the interviewer is to assess why the person is refusing, so that the reason for refusal can be countered.

110. However, the presence of an interviewer may affect the answers given by respondents. If this is felt to be a problem, it may be desirable to arrange for the respondent to interact directly with the computer. This cannot occur if the respondent cannot read or does not know how to operate the computer. The computer can be made very simple to operate, and the interviewer can explain and provide encouragement. The questions can – if sufficiently simple in structure – be recorded as audio files and played to the respondent while they operate the computer.⁴⁸ This requires appreciable effort for a single language and could be difficult to implement in countries which have multiple official languages.

Is a high-quality sampling frame available?

111. Few, if any, sampling frames are without particular features and limitations. The amount of effort (and cost) required to minimise the impact of such limitations is likely to vary from one country to another. If one of the aims is to be able to compare findings across countries, it may be desirable to reduce limitations in the quality of the sampling frame to a greater extent than is usual for single-country studies.

⁴⁶ An incentive of this sort is used universally in in-depth qualitative research, where respondents are being asked to make a significant effort to express their views or beliefs. Generally, in the UK, incentives are in the order of £10 for a one-off face-to-face interview lasting an hour or so. Qualitative researchers usually pay £20 but this sometimes relates to a 1.5 - 2 hours interview. The majority of surveys do not use incentives.

⁴⁷ It is arguable that the younger age group might be more willing to participate, as their education, transition to the labour market and acquisition of skills may be more salient to them than to older people; however, it is unclear whether the relatively low response rate reflects variations in skills or just scepticism about the research having useful outcomes in public policy (or some other points of view).

⁴⁸ This is known as Audio-CASI. The UK National Center for Social Research and BMRB conducted a survey of 12,000 people aged 10 to 65 on crime, in which some parts of the interview used Audio-CASI. Respondents reported favourably on the approach.

Does the research organisation have experience of high-quality coding to international classifications of occupations and industries?

112. Most market research does not require sophisticated (or high-quality) coding and their coding staff may be unfamiliar with international classifications of occupation and industry.

8.2 Funding for the preparatory phases

113. The preparatory phase of PIAAC is expected to last until (up to) the end of 2005. To allow the Secretariat to perform the work until then, adequate resources should be provided through either voluntary contributions or the regular budgets associated with the ELSA or Education committees.

9. Management and governance options

114. PISA, ALL and IALS all provide experience relevant to the effective management of an international programme of skills assessment.

115. In terms of governance, the PISA model could serve as a template for PIAAC. As described in various PISA publications, PISA is a collaborative effort that brings together scientific expertise from the participating countries, steered jointly by their governments on the basis of shared, policy-driven interests. Every country is represented on a Board of Participating Countries. The Board determines, in the context of OECD objectives, the policy priorities for PISA and oversees adherence to these priorities during implementation. This includes the setting of priorities for the development of indicators, for the establishment of the assessment instruments and for the reporting of results.

116. Under PISA, experts from participating countries also serve on working groups that are responsible for linking policy objectives with the best internationally available technical expertise. By participating in these expert groups, countries ensure that the instruments are internationally valid and take account of the cultural and educational contexts in participating countries. They also ensure that the assessment materials have strong measurement properties. Through National Project Managers, participating countries implement PISA at the national level subject to agreed administration procedures. National Project Managers play a vital role in ensuring that the implementation of the survey is of high quality, and verify and evaluate the survey results, analyses, reports and publications.

117. The design and implementation of the surveys, within the framework established by the Board of Participating Countries, is the responsibility of the PISA Consortium, led by the Australian Council for Educational Research (ACER) and including several internationally renowned education testing institutions.

118. The OECD Secretariat has overall managerial responsibility for the programme, monitors its implementation on a day-to-day basis, acts as the secretariat for the Board of Participating Countries, builds consensus among countries and serves as the interlocutor between the Board of Participating Countries and the international consortium charged with the implementation of the activities. The OECD Secretariat also produces the indicators and analyses and prepares the international reports and publications in co-operation with the PISA consortium and in close consultation with Member countries both at the policy level (Board of Participating Countries) and at the level of implementation (National Project Managers).

PART D. THE SECRETARIAT'S STRATEGY PROPOSAL

119. This paper has outlined, in generic terms, the issues informing key strategy choices that need to be made in the development of PIAAC. The experts members of the INES-SMG have recently requested that the Secretariat also propose an overall strategy for PIAAC. This part of the paper outlines the Secretariat's proposal.

Overall character and time-frame

120. PIAAC would be a **multi-cycle programme of assessment that would cover, over time, all the main policy concerns articulated by the IEG**. The first cycle of data collection is envisaged in 2008~9, with reporting in 2010. Subsequent cycles will be administered at five-year intervals. Chart 1 outlines a tentative time-frame for this strategy proposal.

Age-group/policy focus

121. **Each cycle will cover all age groups, but with an age-specific focus or over-sampling.** Age-specific focus might be made on: young adults during the 2008~9 cycle; middle-aged adults during the 2013~14 cycle; and an older cohort of adults for the 2018~19 cycle. However, this can be done in either of two ways:

1. A specific age-group focus in each cycle. We propose to begin with youth, followed by mid-age adults, and then an older cohort (definition of the age range covered by each category can be decided by the IEG). However, each cycle will also include a small sample of individuals from across the age spectrum.
2. A coverage of all age groups in each cycle with over-sampling of youth in the first cycle, middle-age adults in the second cycle, and old-age adults in the third. This means that the sample size has to be large enough. The number of competency domains to cover in this option would need to be decided by the IEG.⁴⁹

122. Having data from across the age spectrum – at least in benchmark form (as in option 1) – is considered necessary in order to explore a number of important policy issues. For instance, data from across the age spectrum could help in examining policy issues that involve, *inter alia*: comparison of competencies across age-groups; assessment of how competencies are acquired and lost over the life cycle; and how the competencies of the entire workforce impact on economy-wide productivity. Whether either option (1) or (2) is selected, **the sample would include the non-employed**. This would allow the examination of a set of policy issues important to both Ministries of Education and Labour. The Secretariat would need to explore how best to assess the competencies of the non-employed.

⁴⁹ In this connection, and as discussed in Part A of this paper, it should be taken into account that an assessment that attempts to cover the entire age spectrum and a wide breadth of competencies might reduce the analytical focus.

Question for IEG members: which of the above two options- 1 or 2 - is preferred ?

Competency domains and assessment methods

123. For the first cycle, PIAAC would directly assess competencies in the domains of literacy, numeracy, problem-solving and ICT. Proven measurement instruments exist for the first three of these domains, and feasibility work on an instrument for the ICT domain has been successfully completed as a part of PISA.

124. At the same time - and subject to validation of the use of the job-requirement approach in an international cross-cultural setting - a survey of job requirements would be carried out, again across all age groups.⁵⁰ The combination of the job-requirement approach with direct assessment would be a feature that distinguishes PIAAC from other international assessments of skills. It would allow the examination, in new ways, of a number of policy issues. For instance, across those competency domains assessed with both techniques, issues relating to possible over-investment in education could be considered by comparing individuals' competencies (assessed through direct tests) with the competencies they actually use in the workplace (assessed through the job-requirement approach). The combination of direct and indirect testing would also afford insight into the merits and demerits of indirect testing technology.

125. The Secretariat proposes that feasibility work begin now to determine how a survey using the job-requirement approach could be implemented as a part of PIAAC. If the results of this feasibility work are positive, and if they are available with sufficient lead time, testing using the job-requirement approach would be included in a first cycle (2008~9). If the results of the feasibility work are positive, but operationalising this approach is likely to require a longer time period, then the implementation of this approach could be scheduled for a second cycle.

126. Earlier in this paper, members of the IEG were asked to indicate the competencies for which there would be an interest in a programme of research aimed at the future development of measurement instruments. Based on the priorities expressed by member countries, research could begin at an early date with a view to fielding instruments in the second and subsequent cycles of PIAAC that would permit the assessment of additional competency domains.

An enterprise survey

127. PIAAC would include an enterprise survey. As described in Part B, such a survey could shed light on a broad array of policy issues not addressed in previous international skills assessments. The Secretariat proposes that feasibility work begin now to determine the viability and *modus operandi* of an enterprise survey. If the results of the feasibility work, as well as the consideration of resource implications, are positive, an enterprise survey could be included in a first assessment cycle. If the results from the feasibility work are positive, but operationalisation is likely to require a longer time period, then the implementation of an enterprise survey could be scheduled for the second cycle. A further consideration in including an enterprise survey is that doing so could provide an opportunity to attract additional financial resources from a variety of institutions other than Ministries of Education and Labour. For instance, Ministries of Industry and Technology, and well-resourced institutions concerned with subnational and enterprise development, would all have a direct interest in the results of such work.

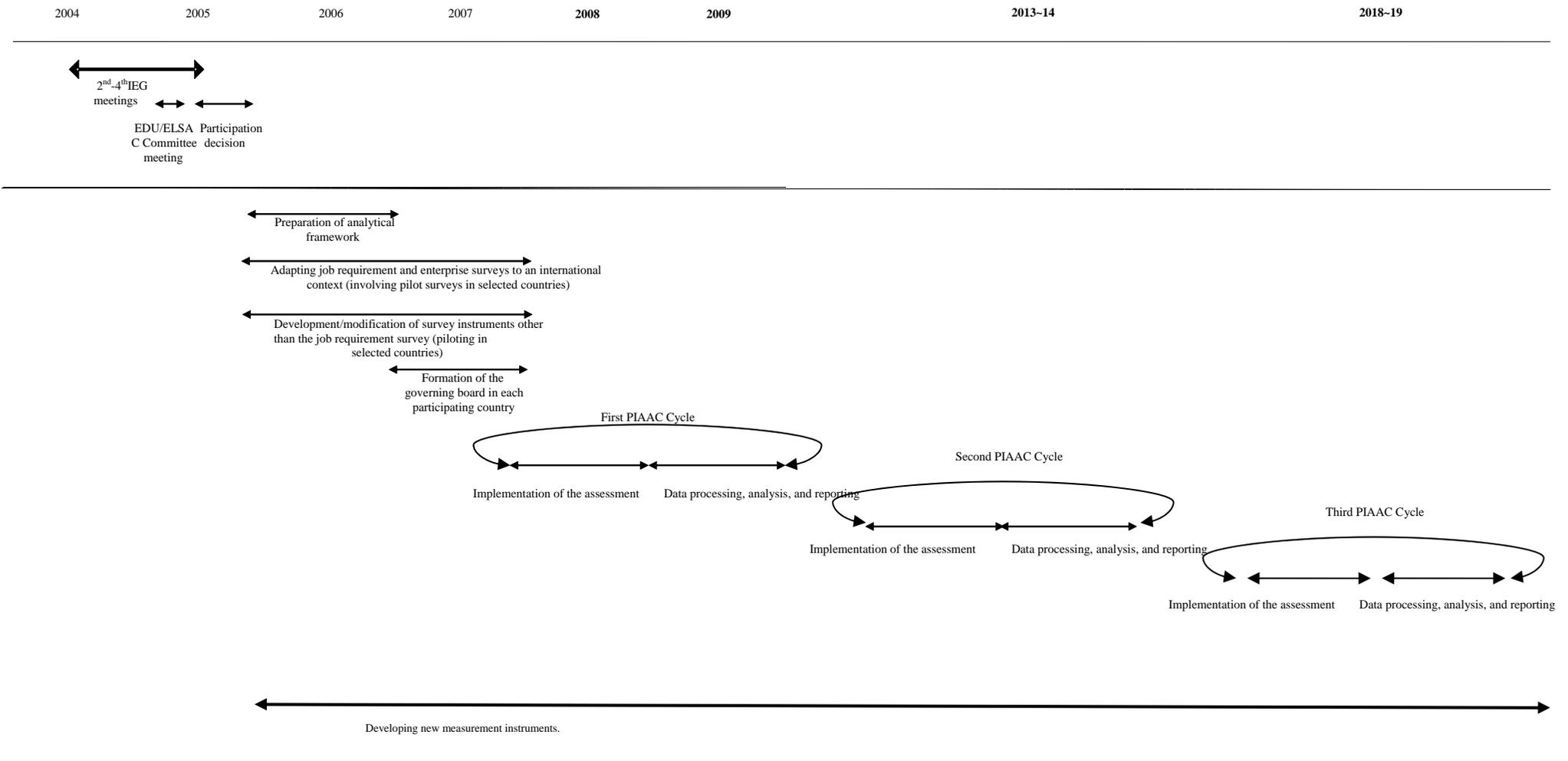
⁵⁰ Some consideration would need to be given to the circumstances under which the job-requirement approach might be used, in combination with direct tests, to assess the competencies of the non-employed.

Other issues

128. Countries could choose to increase the sample sizes and the number of competencies to be assessed if their resources permit. Countries could also choose to introduce a longitudinal design in the data strategy, again if their resources permit. However, while some components of the strategy could be optional, these should be kept to a minimum in order to maximize international comparability.

129. Lastly, countries which have invested in research on skills and competencies may feel they require consistency with these when PIAAC is conducted. The OECD Secretariat has already distributed to IEG members a draft inventory of skills surveys, and requested that IEG members verify the completeness of the inventory. However, if countries wish to maintain consistent questions, this also needs to be identified as early as possible.

Chart 1: A Tentative Time-frame for the PIAAC Programme



REFERENCES

- Abowd, J. M. and H. Finer (1999), Firm and worker heterogeneity and employer-employee matched data in the national longitudinal survey of youth. Manuscript.
- Baker, D.P., Horvath, L., Campion, M. Offermann, L., and Salas, E. (1999), Teamwork: Status Memorandum (draft). Retrieved 5th August 2004 from <http://www.ets.org/all/>.
- Bowles, S. and H. Gintis(2000). "Does Schooling Raise Earnings by Making People Smarter?" in K. Arrow, S. Bowles and S. Durlauf (editors), *Meritocracy and Economic Inequality*, Princeton: Princeton University Press.
- Bradley, S. and Nguyen, A.N. (2004), The School to Work Transition, Chapter 14 in, *International Handbook of Education Economics*, forthcoming, Edward Elgar.
- Coulombe, S., Tremblay, J-F., and Marchand, S. (2004), Literacy Scores, Human Capital and Growth Across Fourteen OECD Countries, Statistics Canada and Human Resources and Skills Development Canada. Ottawa.
- Crouch, K.A. (1992), New evidence on the long-term effects of employment training programs, *Journal of Labor Economics*, 10(4), pp.380-88.
- De la Fuente, A. and A. Ciccone (2003), *Human Capital in a Global and Knowledge-based Economy*, European Commission, DG for Employment and Social Affairs, Office for official publications of the European Communities, Luxembourg.
- Dolton, P.J. (2004), The economic assessment of training schemes, Chapter 3 in, *International Handbook of Education Economics*, forthcoming, Edward Elgar.
- Dolton, P.J. and Makepeace, G.H. (2004), Use IT or lost IT? The impact of computers on earnings, University of Cardiff, mimeo.
- Economist Intelligence Unit (2004), Reaping the benefits of ICT: Europe's productivity challenge, London.
- Felstead, A., Gallie, D. and Green, F. (2002), Work Skills in Britain 1986-2001.
- Glaeser, E.L. and Maré, D.C. (2001), Cities and Skills, *Journal of Labor Economics*, Vol.19, No.2 (April).
- Goldsmith, A.H., Veum, J.R. and Darity, W. (1997), The Impact of Psychological and Human Capital on Wages, *Economic Inquiry*, Vol.XXXV, October 1997, pp.815-829.
- Gradstein, M. and Justman, M. (2002), Education, Social Cohesion, and Economic Growth, *The American Economic Review*, Vol.92, No.4, September.
- Green, D.A. (2001), Literacy Skills, Non-Cognitive Skills and Earnings: An Economist's Perspective, http://qed.econ.queensu.ca/pub/jdi/deutsch/edu_conf/Green.pdf

- Green, F., S. Machin and D. Wilkenson (1998). The Meaning and Determinants of Skill Shortages, *Oxford Bulletin of Economics and Statistics*, Vol. 60, no. 2, May, pp. 165-88.
- Green, F.S., Ashton, D. and Felstead, A. (2001), Estimating the Determinants of supply of computing, problem-solving, communication, social, and teamworking skills, *Oxford Economic Papers* 3, pp.406-433.
- Green, F. (2004), First thoughts on methodological issues in an international assessment of adult skills, Paris, OECD. (COM/DELSA/EDU(2004)6).
- Heckman, J. and Vytlačil, E. (2000), Identifying the role of cognitive ability in explaining the level of change in the return to schooling. NBER Working Paper series 7820. Cambridge, MA: The National Bureau of Economic Research.
- Hellerstein, J. K. and Neumark, D. (2004), Production function and wage equation estimation with heterogeneous labor: evidence from a new matched employer-employee data set. NBER Working Paper Series 10325. Cambridge MA. The National Bureau of Economic Research.
- Knack, S. and Keefer, P. (1997), Does Social Capital Have an Economic Payoff? A Cross-Country Investigation, *Quarterly Journal of Economics*, Vol. 112(4), pp. 1251-1288.
- Krugman, P. (1994), Competitiveness: A Dangerous Obsession, *Foreign Affairs*, March/April.
- Lennon, M., Kirsch, I., Von Davier, M., Wagner, M and Yamamoto, K. (2003), Feasibility Study for the PISA ICT Literacy Assessment, Australian Council for Educational Research (ACER), National Institute for Educational Policy Research (NIER), Educational Testing Service (ETS). See http://www.pisa.oecd.org/Docs/Download/ICT_Feasibility_Report.pdf
- Mariachristina, P., Santarelli, E. and Vivarelli, M. (2003), The Skill Bias Effect of Technological and Organisational Change: Evidence and Policy Implications, IZA Discussion Paper Series, Discussion Paper No.934, November 2003.
- Murray, S.T. (2003), Reflections on international competence assessments, in Swiss Federal Statistical Office (2003).
- National Audit Office, United Kingdom (2004), Welfare to Work: Tackling the Barriers to the Employment of Older People, Report by the Comptroller and Auditor General, London, September 2004.
- OECD (2001), *The Well-Being of Nations: the Role of Human and Social Capital*, Paris.
- OECD (2000), *Literacy in the Information Age: Final Report of the International Adult Literacy Survey*, OECD/Statistics Canada.
- OECD (2002), *ICT Skills and Employment*, STI Working Papers, Paris
- OECD (2003), *The Sources of Economic Growth in OECD Countries*, Paris.

- OECD (2003a), The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving, Knowledge and Skills, Paris.
- OECD (2004), LFS database.
- Putnam, R. (2000), *Bowling Alone: The Collapse and Revival of American Community*, Simon Schuster, New York.
- Putnam, R. (2001), Social Capital: Measurement and Consequences, in J.F. Helliwell (ed.), *The Contribution of Human and Social Capital to Sustained Economic Growth and Well-being: International Symposium Report*, Human Resources Development Canada and OECD.
- Ritzen, J. and Woolcock, M. (2000), Social Cohesion, Public Policy, and Economic Growth: Implications for Countries in Transition, Address prepared for the Annual Bank Conference on Development Economics (Europe), Paris, June 26-28, 2000.
- Rudd, R.E., Kirsch, I. and Yamamoto, K. (2004), Literacy and Health in America, Centre for Global Assessment, Policy Information Center, Research and Development, Educational Testing Service, Princeton, New Jersey.
- Rudd, R.E., Moeykens, B.A. and Colton, T.C. (1999), Health and Literacy: A Review of Medical and Public Health Literature, in *Annual Review of Adult Learning and Literacy*, Comings, J., Garners, B. and Smith, C. (eds), New York: Jossey-Bass, 1999.
- Rychen, S. and Salganik, L.H. (editors) (2003), *Key Competencies: for a Successful life and a Well-Functioning Society*, Göttingen, Germany: Hogrefe & Huber.
- Sianesi, B. and J. Van Reenan (2003), The Returns to Education: Macroeconomics, *The Journal of Economic Surveys*, Vol. 17, No. 2, pp. 157-200.
- Swiss Federal Statistical Office (2003), Contributions to the Second DeSeCo Symposium, Geneva Switzerland, 11-13 February, 2002. Neuchâtel.

APPENDIX A. EXISTING INTERNATIONAL COMPARATIVE ASSESSMENTS OF SKILLS

A.1 International Adult Literacy Survey (IALS)

130. The International Adult Literacy Survey (IALS) was the first international household-based assessment of adult literacy skills. The result of a collaborative endeavour between the OECD, Statistics Canada, the US National Center for Education Statistics and the US Education Testing Service, IALS was administered to a nationally representative sample of adults aged between 16 and 65. The survey was administered in three cycles of data collection between 1994 and 1998. The 1994 cycle included nine countries – Canada (English and French-speaking populations), France⁵¹, Germany, Ireland, the Netherlands, Poland, Sweden, Switzerland (German and French-speaking regions) and the United States. Five additional countries or territories administered the IALS instruments in 1996: Australia, the Flemish Community in Belgium, Great Britain, New Zealand and Northern Ireland. Nine other countries or regions participated in a third round of data collection in 1998: Chile, the Czech Republic, Denmark, Finland, Hungary, Italy, Norway, Slovenia and the Italian-speaking region of Switzerland.

131. IALS measured three domains of literacy: prose, document and quantitative. Prose literacy refers to the knowledge and skills needed to understand and use information from texts, such as newspaper articles and fiction. Tasks in this domain included inferring a theme from a poem, finding an item of information in a newspaper article, and interpreting instructions in a guarantee.⁵² Document literacy refers to the knowledge and skills required to locate and use information contained in a wide range of document types, such as maps, tables, graphs and schedules. Tasks in this domain included locating a junction on a street map, using a timetable to select an appropriate bus, and completing an application form. Quantitative literacy refers to the knowledge and skills required to apply arithmetic operations using numbers in printed materials. Tasks in this domain include estimating interest payments from a loan advertisement, calculating a tip and balancing a checkbook. Respondents were also asked to report on: how frequently they engaged in literacy activities at work with different types of texts; writing and reading practices at home; and how adequate they considered their literacy skills were.

132. IALS illustrated the critical role of education and learning for individuals, throughout life, and for the economy as a whole by providing insights into the distribution of skills across demographic and social groups. Analysis of the data generated by IALS has shed light on the role of literacy skills in macro-economic performance,⁵³ as well as better understanding of how inadequate literacy skills hinder

⁵¹ France eventually withdrew from the survey.

⁵² In IALS, three characteristics were used in the creation of tasks: contexts/content, materials/texts, and processes/strategies. Selecting a wide range of contexts for the tasks was aimed at helping to avoid biases for or against any particular group of adults. Six content categories were identified: Home and family; health and safety; community and citizenship; consumer economics; work; and leisure and recreation. Materials used included continuous – or prose - texts and non-continuous texts. The processes employed refers to the way in which test takers respond correctly to a question or directive.

⁵³ See Coulombe *et al* (2004).

economic, political and social participation. IALS has also afforded an analysis of linkages between literacy and a range of outcomes in OECD labour markets. And IALS has provided the finding that in all countries older age cohorts experience an important skills deficit with respect to younger age groups. Furthermore, IALS has involved important methodological advances in international comparative assessment.

A.2 Adult Literacy and Life Skills Survey (ALL)

133. The Adult Literacy and Life Skills Survey (ALL) aims to measure the distribution of a range of skills considered important to economic and social success in a nationally representative sample of adults between 16 and 65 years of age. ALL is a household survey. ALL is sponsored by Statistics Canada, the U.S. National Center for Education Statistics, a regional UNESCO office and the OECD. Seven countries fielded the ALL pilot and main studies between 2002 and 2004: Bermuda, Canada, Italy, Mexico (Nuevo Leon), Norway, Switzerland and the USA. Other countries will participate in a second round of data collection between 2004 and 2006. The first round of ALL results will be published in the first half of 2005.

134. ALL uses identical test items to IALS in the measurement of prose and document literacy. However, ALL broadened the IALS numeracy domain. Participating countries also have an option to include a fourth domain covering analytical problem-solving. It was intended that ALL include tests to measure teamwork, practical knowledge and perceived familiarity with and use of new information and communication technologies (ICT). However, independent studies concluded that the approaches to measurement of teamwork and practical cognition would not afford data of the required quality. ICT skills were examined indirectly through self-assessment as part of the background questionnaire.

135. ALL is expected to provide data and analyses relevant to the following sorts of question: What is the distribution of literacy skills among adults in the participating countries? What is the relationship between these literacy skills and the economic, social, and personal characteristics of individuals? What is the relationship between these skills and the economic and social characteristics of nations? To what extent do adults perceive that they are familiar with information and communication technologies?

136. Respondents in the ALL survey completed a background questionnaire that included questions on the respondent's: educational attainment; linguistic background; parental background (including schooling and occupation); labour force activities (including time spent employed, unemployed or in job-search; occupation and type of work; employer characteristics; work-based remuneration); literacy and numeracy practices at work; participation in education and learning; general literacy and numeracy practices; practices related to social capital and physical well-being (including the frequency of use of a library or bookstore; time spent watching television or videos; newspaper reading practices; volunteering; and health status); personal use of a range of technologies (including purpose and type of use of computers and the Internet; perceived adequacy of computer skills); household information and income.

APPENDIX B. ADAPTING THE JOB-REQUIREMENT APPROACH FOR AN INTERNATIONAL SURVEY

137. A key issue is how readily the use of the job-requirement approach might be extrapolated to an international context. As described in the main body of this text, Green (2004) suggests that several types of development work are required. This appendix elaborates on the procedures, possible resource implications, and proposed schedules associated with these steps.

The issue

138. The job-reporting method asks about a large range of identified activities, and elicits responses on a 5-point scale of importance. The main problem to investigate is the cross-national validity of the importance scale. Even after a high-quality translation exercise, respondents in different countries might assign different meanings to the points on the importance scale, such as “very important” or “essential”. *A priori*, it is not known whether this is a serious problem. The pilot survey proposed below is designed to collect data to test this. A secondary problem concerns the identification of activities. The activities described in the survey are, intentionally, generic. Lacking particularity, respondents from different countries might assign different meanings to vaguely-worded activities. Some validation checks are therefore required here also. This will be examined using cognitive research methods.

Solutions

139. Two kinds of validity check on the importance scale are proposed.

- i) For three activities – use of physical strength, use of computers and writing long documents – additional questions can be asked to check in greater detail how these activities relate to the jobs. These can include questions asking about how crucial these activities are to the job, the frequency of the activity (which in some cases, though not all, is expected to be linked to importance), and possibly also the horizon.⁵⁴ The idea is to investigate if international comparisons of these particular skills are

⁵⁴ Examples of draft alternative, or related, importance questions are:

Cruciality: “Suppose that, for some reason, you became permanently unable to carry, push or pull heavy objects, how would this affect your ability to carry out your job ?”

Scale: No effect / “Able to carry on, but would not be able to do the job properly” / “Could not continue doing the job”

Frequency: “How often does your job require you to carry, push or pull heavy objects?”

Horizon: “Suppose that your computer or computerised equipment broke down and none other were available. For how long could you carry on doing your job productively?”. Scale: No time – would have to stop immediately/ 1 hour/ half a day/ etc. etc.

consistent across the multiple measures available.⁵⁵ This statistical method of validity checking requires that samples be large enough to estimate population means with reasonably small standard errors. The suggested sample of 500 in each country would be adequate. In addition it is proposed to over-sample a particular occupation (e.g. teachers). This will enable a statistical within-occupation international check along the same lines, but with the advantage that the standard errors of the mean skill levels would be lower in each country, so that a more precise validity check is facilitated for this particular group. 100 would be needed within the chosen occupation to be included in the sample of 500.

- ii) As a second validity check, it is proposed to convene an expert group drawn from each of the participating countries to examine and investigate the findings concerning skill differences within the particular occupation. The objective will be to examine whether the differences in the skills for this particular occupation, with respect to the three skills identified, can be explained by experts' prior knowledge, in a shared panel discussion. This method will be particularly relevant if the above statistical check reveals anomalies that might otherwise cast doubt on the validity of the comparisons.

140. As a way of testing the validity of the identification of activities, it is proposed that, in respect of activities where there is some doubt as to whether the meaning is the same in different countries, a pre-pilot cognitive test is carried out in each country.

The Purpose of Cognitive testing

141. Cognitive testing aims to understand the cognitive process the respondent has to accomplish in answering the questions. It is a qualitative research process, meaning that it involves small samples and intensive interviewing of each sample member. Among a range of techniques that may be employed, a key method is to ask a question and record the answer given. The respondent may then be asked a series of further questions about how they interpreted the question and how they decided to answer it: what did they understand certain words to mean ?; was there any part of the question that was unclear to them (at first hearing) ?; how sure were they about the accuracy and completeness of the answer they gave; in what respects were they unsure and why, for example was the problem one of memory, sensitivity/embarrassment, etc ?; having considered the question at greater length, would they wish to change their answer ?

⁵⁵ The method could be succinctly stated as follows: let S_{1A} be the average measure of the first skill in country A (or of a sub-group in A); similarly S_{1B} in country B.

Check 1: if $S_{1A} = S_{1B}$ according to the importance scale, is this also the case according to the other importance measures? Do this check for all three activities; cover different parts of the skills spectrum.

Check 2: if $S_{1A} > S_{1B}$ according to the importance scale, is this also the case according to the other importance measures? Is the difference $(S_{1A} - S_{1B})$ correlated across measures? Do this for all three activities; cover different parts of the skills spectrum.

Taken together, this is already quite a strong set of tests of international validity. An additional statistical check could also be made regarding the ranking of the variance of the skills estimates for the three chosen activities. Consideration will need to be given to exactly how crucial each test is. At what point is it decided that the validity is inadequate to proceed with inclusion of job-reporting of generic skills in PIAAC?

142. A number of interviewers are needed to test the questions in different communities and regions. Each interviewer may test a number of versions of the questions with different respondents. This enables the gathering of a variety of contrasting views about how well the questions work when different forms of wording are used with people of different backgrounds. While there is a temptation to interview articulate respondents for this exercise, it is important to make sure that less well-educated people are included in the sample. An approach that is seldom used, but which may have some value in the context of a major cross-national study, is to re-interview some of the sample members when they have had time to think about the exercise, to see whether they give consistent answers when interviewed a second time.

143. There may be an argument for this sort of in-depth exercise to be repeated in every country, since part of what it usually does is to enable researchers to anticipate where problems of interpretation may arise. It does not necessarily offer solutions to problems, and in some cases it may be necessary to accept that standardisation across countries cannot be achieved at the same time as a sufficient level of comprehension at the level of the 'lowest common denominator' in all samples. This indicates that it is necessary to devise methods to 'triangulate' assessments from a number of different directions. There is no substitute for actual data to enable this to be done, and this implies that a small number of countries should conduct large-scale pilot studies in order to generate sufficient data for the methods of analysis to be worked out in advance of the major investment of conducting a standardised study across many countries.

B.1 Proposed procedure

- i) Re-draft the questionnaire in English to :(a) focus on the generic and/or specific skills to be covered by PIAAC; (b) adapt questions to suit an international comparison.

144. At this stage, it would also be necessary to identify the further questions to be used as background information on respondents and other interpretative information.

145. Green (2004) suggests that the job-reporting questions need to re-cast from asking about the 'importance' of competencies to asking about the frequency with which they are used as part of the individual's job. In the 2001 UK Skill survey, the answer scale for 'importance' had the following five categories: essential; very important; fairly important; not very important; not at all important; and does not apply. The use of such a standard set of categories was conducive to self-completion on the computer in two ways. Firstly, it made the task simple, once the respondent had been shown how to input their answer for the first one or two questions. Second, it avoided the interviewer having to read out a long list of similar-sounding questions.

146. It is unlikely that it would be helpful to increase the number of answer categories. The issue is how to decide on a series of answers that relate to frequency and that would be applicable in virtually all work situations. Issues may arise in situations such as shift-work and part-time working, where the time spent in work may vary from one time-period to another. Some competencies may be used very infrequently.

- ii) Constitute an 'expert panel' to consider the first draft of the questions.

147. As far as possible, the researchers undertaking the development work need to anticipate issues that will arise for different respondents. An efficient way to do this is an 'expert panel', consisting of interested researchers and others who are knowledgeable about how work is organised, such as occupational psychologists and human resources specialists. They are asked to consider a draft set of questions, and then meet to exchange ideas about the problems that may arise and make suggestions about how these problems can be minimised or avoided. This is a fairly low-cost process, and is likely to be cost-effective.

- iii) Engage a suitably qualified survey company in each of three countries, and meet to examine the English questionnaire.
- iv) Translate the questionnaire (using double translation followed by reconciliation, that is, with three checks into the target language).
- v) Conduct face-to-face interviews in each of the three countries, using the same questionnaire. However, to assess cross-country and cross-cultural validity an analysis in a cognitive laboratory in each country may be required (a mix of frequency and importance scales for measuring activities can be tested during this pilot phase).

148. It may be advisable to undertake a two-stage cognitive test of key components of the survey design (along the lines of the cognitive testing conducted for the 2001 UK Skills Survey⁵⁶). Having a second stage would allow re-testing where the first stage has identified problems, to make sure that the solutions devised are going to work better than the original versions.

149. It is suggested that, at each stage of the cognitive testing, five interviewers⁵⁷ are employed to conduct between six and eight interviews over a period of around ten days (the size of the sample would depend to some extent on the age range to be covered in the main survey, and might not need to be greater than 30-40 interviews at each stage). The interviewers attend a face-to-face briefing meeting before starting work and a similar debriefing soon after completing their work. Each interview lasts around 60 minutes and respondents would be given a cash incentive (in Britain, currently around 15-20 Euro). Interviews should be recorded for review by researchers, with the permission of respondents. As the number of questions that can be explored in each interview is fairly limited, it is suggested that an interview plan is drawn up that systematically allocates questions to respondents. This may include alternative versions of some questions. To ensure that a variety of work situations are covered, each interviewer would have a quota to achieve, specifying for example occupational levels and details such as part-time working. As a result, interviewers may need to spend an appreciable amount of time finding suitable people to interview. It is not envisaged that such interviews need be conducted using computers, since the issues are about wording rather than the mode of interview. It is probably easier for interviewers to note their observations on a paper questionnaire than in a computer.

- vi) Conduct a further stage of consultation involving other countries in which cognitive testing had been conducted, aiming to resolve major issues of vocabulary.
- vii) Conduct a 'pilot survey' of around 500 interviews, based on a probability sample, which would involve development of the full repertoire of survey materials and their implementation. This would generate a dataset (to be combined with those of 2 to 3 other countries).

150. In terms of costs, for each stage of the work the key elements are:

- Ten interviewer-days to attend the briefing and debriefing, including travel expenses (and perhaps the costs of venues for the meetings and refreshments for those participating in them).

⁵⁶ The 2001 Skills Survey cognitive test explored a suitable title for the study as well as covering a varied set of 12 questions, which were divided into sets of six questions put to each respondent.

⁵⁷ The interviewers need to have been given special training in the in-depth interview techniques required for cognitive testing; however, it is probably helpful for them also to conduct structured interviewing, since the outcome of the test is better questions for use in structured interviews.

- Twenty interviewer-days to conduct interviews (with additional time to screen a set of addresses for eligible people, if the quota means they need to contact a large number of people to find the right people to interview).
- Incentives for respondents.
- Transcription of interviews (optional).
- Researcher time to draft questions and briefing materials, produce materials for use in the interviews (e.g. lists of answers to be shown to respondents).
- Researcher time to analyse the findings and prepare a report on the cognitive test.

151. It can be seen that the overall costs of the exercise are fairly modest. With a two-stage cognitive test, the interviewers will not need to attend a briefing for the second part, unless the nature of the questions has changed appreciably.

152. The amount of researcher time required is not easy to specify in advance. Clearly, the cognitive test is an important factor in overall survey quality, and so it warrants a reasonable amount of time. However, the time required may depend on factors such as how far the commissioning organisation has anticipated which questions require testing; if this has not occurred, then a fair amount of thought may be required to identify the most useful set of topics for inclusion in the test. The analysis of the interview material needs to be done systematically, working through the recording of each interview in turn to identify additional details that the interviewer may not have mentioned in the debriefing. Around eight to ten researcher-days is probably the minimum for a typical cognitive test of this scale.

153. It was suggested above that as part of the initial development process, a small number of countries should conduct a large-scale initial survey, to generate data to allow the analyses for the main stage study to be developed and checked. This would be a relatively costly exercise, but is probably justified in the context of a major international study.

B.2 Resource implications

154. **The following cost estimate is based on an assumption that this pilot will be held in the United Kingdom.** The approximate estimate which covers both the cognitive testing stage and a pilot survey of 500 cases would be in the region of 94,000 EUR in Britain. The main elements are as follows⁵⁸:

Researcher time and expenses:	22,000 EUR
Cognitive interviewing:	6,000 EUR
Pilot survey (interviewers and field management)	66,000 EUR

⁵⁸ Note: the actual costs would be dependent on the details of the specification for the work (e.g. they might change if the selected occupational group was less numerous than teachers or if other parts of the questionnaire were to be tested).

155. Note that this estimate does not include fees for consultants. Total required work days are estimated to be around 72. Including administrative support, consultant fees may amount to around £59,000 (84,500 Euros). Nor does this estimate include translation costs.

B.3 Proposed schedules

Activities	Timeline
0. Identify countries to take part in the pilot, including funding the pilot survey in their own country, according to the specifications below. It is proposed that the three countries be: an English-speaking country, a continental European country, and an Asian country.	Nov~Dec, 2004
1. Research team to re-draft the 2001 Skills Survey questionnaire to generate a preliminary English-language version of the pilot questionnaire for the international pilot.	Nov~Dec, 2004
2. Survey experts' panel meeting to: <ul style="list-style-type: none"> • express and discuss possible initial problems with the wording of the questionnaire • review procedures for the overall management and conduct of the project 	Jan~Feb, 2005
3. Research team redrafts questionnaire ready for translation. The two non-English speaking countries arrange translation. Back translation, checking and subsequent iterations, are subsequently arranged by the research team.	March, 2005.
4. Cognitive testing/re-testing and pre-pilot preparation. Re-drafting and re-translation arranged as necessary. Member(s) of research team to visit national survey teams <i>in situ</i> , prior to final design of questionnaire and start of fieldwork.	April, 2005.
5. Conduct of a pilot survey in each of three countries. Preparation of pilot data sets in either SPSS or STATA format.	May~June, 2005.
6. Analysis of findings by research team.	August 2005.
7. Experts' panel meets to review comparisons, and discuss anomalies.	Early September 2005.
8. Report prepared for OECD, prior for presentation as required to PIAAC committee of experts.	End of September 2005.

APPENDIX C. CLASSIFICATION OF SURVEY COSTS

156. It is assumed here that in each country there will be a team of researchers and an organisation conducting the data collection and preparation of a dataset. These persons may be members of the same organisation, or not. The costs defined here relate to the activities of the organisation collecting the data. It is assumed that this organisation has the necessary expertise to select a high-quality sample and has the capability to provide estimates of complex standard errors and design effects (or that they will contract these from suitable specialist organisations, so that these costs are part of their contract to undertake data collection). These costs are treated as part of the cost of running the survey, although they could be incurred by the 'research team' in some countries (e.g. if an academic group has particular strengths in these areas).

Survey administration

157. A large part of the costs will be those involved in the survey organisation's day-to-day operations, including its premises, its staff and IT systems. These are likely to be reflected in 'overhead costs', which will be spread across the range of surveys that are on-going at any point in time. Where a national statistical organisation is contracted to conduct data collection, they may or may not treat overheads as part of the budget for running a specific study. If they do not do so, they are likely to be less costly than other organisations (such as commercial or other independent organisations) that do.

Sampling

158. This is fundamental to a study and may require significant expenditure. As noted above, the market research industry is, on the whole, weak on sampling and organisations may be reluctant to agree that their usual quota or quasi-random 'sampling' methods are unsuited to a study of this sort.

159. Programming the questionnaire

160. Different computer-assisted personal interviewing (CAPI) systems have varying ways of implementing the design of questionnaires. There is a logical distinction between:

- declaring the wording of the questions and the answer categories, and
- defining the rules that determine which questions are asked of which respondents and in what order.

161. Even if the wording of questions has been agreed by all countries, there is potential for different interpretations of the 'filter rules'. Different software systems implement the rules in different ways, giving considerable scope for the final data to differ between organisations. As a result, it is necessary for an agreed set of scenarios (examples of respondents' answers to the questions) to be input on each country's own system. As this is being done, the sequence of questions is recorded, to make sure that the programming of the 'rules' is working in exactly the same way as it does on a reference software system.

162. Organisations may have the implementation of the CAPI questionnaire carried out by researchers or by specialist staff ('programmers'). In either case, this work must be done prior to piloting and requires a substantial investment of time. Most organisations will not have had experience of the amount of effort needed to achieve consistency when different software systems are being used, and they may tend to under-estimate the cost of this work. Of course, if several countries are going to use the same software package, it makes sense for them to exchange the program code to ensure consistency.

Briefing of interviewers

163. It is assumed that a fieldforce and a management system to administer fieldwork, conduct quality control, etc is already in existence. The subject-matter of a survey on adult competencies is sufficiently specialised to warrant at least a one-day training session for every interviewer.⁵⁹ This would cover, at a minimum: the selection of respondents (e.g. criteria by age, working status, etc); background information on the measurement of adult competencies (partly so that the interviewer can explain why the study is important); a worked example of a complete interview; more detailed examination of specific types of question; anticipation of the types of question likely to be raised by respondents and problems that respondents may wish to refer to the interviewer; the timetable of the study; whom the interviewers should contact about their progress and any problems they may encounter.

Conducting data collection

164. Maximising the response rate means ensuring that numerous opportunities arise for an interviewer to contact each sample member. Contacting people in households where all adults are in paid work may be especially demanding, both in finding them at home and in making an appointment to conduct the interview. People living alone may be especially difficult to contact. It is necessary for interviewers to be available to work when most employees are not in work (e.g. in Britain this means at evenings and weekends).⁶⁰

165. One way the amount of work required may be estimated is to consider a typical working day for an interviewer of six hours (in Britain, this would be mid-afternoon to late evening, say 3.00pm to 9.00pm). If an interviewer is lucky and has been able to set up appointments, they may achieve three interviews in this time. This implies that they have three hours within the sample area to get from one address to another and to make contact with other members of their sample. However, some respondents are slow; they may take 80 minutes or more to do an interview that most people do in 60 minutes. This means the interviewer usually needs to allow 90 minutes between appointments (as their relationship to respondents means they have to be punctual). As a result, an interviewer might, realistically, manage an average of between 2 and 2.5 interviews in a 6-hour day. This allows time for contacting people who need 10 or more attempts to find them at home⁶¹, travelling between the addresses and locating the addresses 'on

⁵⁹ This is one reason why it is necessary for each interviewer to undertake a reasonably large amount of work on the project.

⁶⁰ An interviewer who is reluctant to work in evenings and weekends is likely to achieve less work per day than someone who works efficiently. Most organisations employing interviewers in Britain use a system of payment for achieved interviews, with a smaller amount relating to time spent on the work.

⁶¹ Most organisations specify a minimum number of 'calls' that an interviewer must make before recording the outcome as 'non contact', often four or five. These calls must be made on different days and at different times of day (this is not to stop an interviewer from making several calls on the same day, but just to treat these as one call in terms of the minimum standard). Some organisations encourage interviewers to keep making calls after reaching the minimum, as long as they believe there is the prospect of an interview being achieved.

the ground'. Hence, if the target is for each interviewer to achieve 20 interviews on average, this assignment is likely to require 8 to 10 days (most interviewers work 2, 3 or 4 days per week).

166. If the sample design involves a significant element of screening, then time for this needs to be included, having the effect of reducing the 'rate per day'. As the incidence of eligible sample members may vary from one area to another, it can be unfair to base the payment for this work solely on the interviewer's productivity.

Pilot survey work

167. Quite apart from pre-testing of questions, a 'dress rehearsal' pilot survey is required before the main stage of data collection. This is concerned with ensuring that all systems are operating effectively, albeit on a modest scale (say, 50 to 60 interviews). It allows the survey administration procedures to be fine-tuned to maximise the cost-efficiency of the main survey, and hence should be considered essential, even where changes to questions are not expected to be required.

Data preparation

168. Computer-assisted interviewing has led to the virtual elimination of missed data due to interviewer errors. It also allows some types of inconsistency in the data to be checked with the respondent during the interview, subject to the interviewer having sufficient knowledge or skill to be able to resolve the problem. The program may identify details such as extreme values on financial data, to avoid interviewers mis-placing the decimal point when recording money sums. Nevertheless, complex data is almost certain to require some intervention by skilled coders, to code parts of the data and to edit it where appropriate.

169. Coding is required for open questions and for occupation and industry coding (unless the interviewer has been trained to do these tasks, subsequent to the interview). At the UK National Center for Social Research, much of this coding is based on printed reports, and the data are incorporated into the CAPI data file as part of a later editing stage. This includes a check on 'other answers' at partly-precoded questions, to see whether they are substantially the same as a precoded category of answers.

170. Editing involves the coder checking certain details of the data. In some cases, an interviewer may have made notes to explain how the respondent interpreted certain questions. In other cases, the coder may examine unexpected patterns of answers, although on the whole it requires the researchers' view as to whether or not to accept such data. In extreme cases, a record may be deleted from the data file.

171. A further aspect of data preparation is the labelling of variables and values in a data file and (according to a researcher's preferences) declaration of missing values. Quality checks are required on the final data file, especially where the data has a complex structure. In some cases, more than one data file will be output. Documentation is required to enable users to make proper use of the data.

Data weighting

172. The final data file needs weights for use in the analysis. It may need a number of weights to provide for different aspects of the analysis. If certain types of respondent have been over-or under-sampled, they may require different weights to make the overall sample representative of the population.

173. Scaling or grossing weights may be used to enable survey data to represent estimates for nations, regions or other entities. On the whole, the standard approach at the National Centre for Social Research is to make the weighted sample equal to the number of interviews. This ensures that statistical tests are not misled by an apparently larger or smaller sample size. The unweighted base is quoted in every analysis.

Preparation of a technical report

174. The technical report provides the user of the data with information they need to assess the quality of the survey as well as details such as the dates of data collection, the exact sampling frame and how it was applied. It may include tables which compare the survey findings with other estimates of the same phenomena. For example, the survey may have included details such as the type of accommodation or level of household income. These can usually be compared with other statistical sources and may suggest the direction and magnitude of bias in the survey's estimates.

Project management

175. The organisation handling the study may have a number of staff with different job titles who contribute to the handling of the survey. The key staff are 'researchers' who have responsibility for implementing the survey specification with the standard methods used by their organisation. They coordinate the various departments of the organisation and conduct briefings of interviewers and of coding staff.

Classifying fixed and variable costs

176. The costs outlined above can be broadly categorised into fixed and variable costs. Most headings incorporate both of these elements. If we assume a fairly large survey, the research staffing costs involved in project management are likely to contain both in fairly equal proportions. For example the project setup and pilot survey may be essentially fixed costs, while management of data collection, data preparation and the technical report may, to some extent, be scale-dependent.

177. Some survey-specific interviewer training costs may be essentially fixed, but on the whole most data collection and data preparation costs are variable. If the budget is limited, there may be merit in extending data collection over a long period of time (six months or more), so that each interviewer who is trained can conduct two or more assignments, each in a different area. This may have benefits in their gaining expertise in handling the specific requirements of the survey (but, on the whole, interviewers may prefer to work on a varied range of studies, especially when it is not easy to achieve a high response rate, as is likely on this study).

178. As with most areas of business activity, the costs quoted by survey organisations are likely to have a direct connection to the quality of the services they provide. There are only limited opportunities for real efficiency gains to be made between a number of organisations that are offering the same services. Hence, if one quotation is especially low, it is likely that the organisation concerned has under-estimated the work required or has simply made an error in its estimates; in either case, they are unlikely to perform work for which they are not being paid and the result will be a reduction in quality in some part of the process.

179. Survey quality has many dimensions. The response rate is often quoted as an indicator of quality, and this is because non-response has a detrimental effect across the aspects of a study's validity. However, the quality of a dataset depends on levels of item (non) response, which may be affected in various subtle ways by the respondent's perception of the task they are being required to undertake. For example, unless an interviewer presents the interview task as serious, it is unlikely that many respondents will do so.

APPENDIX D. IMPLICATIONS OF USING A COMPUTER TO ADMINISTER SURVEYS

180. It has been assumed that the survey interview will be administered by the interviewer. Other parts of the interview (perhaps up to half) may be self-administered by the respondent using a computer, with the interviewer still in attendance. The interviewer can be instructed to 'train' the respondent in the software during the early part of the interview. As noted above, the task of self-completion can be made reasonably simple, typically involving use of several number-keys to indicate the answer(s), with the 'Enter' key used to move on to the next question. Aspects of screen design can help to keep the task simple. On the whole, respondents are eager to use the computer, but interviewers need to be alert to any reluctance or inability to do so. In this situation, the interviewer can ask the questions in the conventional way.

181. In the United Kingdom – to take one country - it can be almost taken for granted that people aged up to 25 (and most aged up to 35) will be familiar with computers. Even if computers were not in use when some of the older sample members were at school, they are likely to use some sort of microprocessor-based equipment in their work. However, even now, it is still necessary to offer the option of the interviewer asking questions in the conventional way and entering the answers on the computer. This is necessary, for instance, when the respondent has impaired vision. This approach generally works well. Its main limitation is when the respondent is reluctant to answer honestly, for example if the true answer would be socially undesirable (e.g. shows them to have limited skills). How they answer sensitive questions may also be affected by the presence of other family members.

182. There are a number of advantages in the respondents completing part of the interview on the interviewer's computer. For example, this allows them to work at a pace they find comfortable. It is believed that one of the advantages of Audio-CASI may be that it encourages respondents to listen to the whole question, when they may tend sometimes to read only as much as they think they need to work out their answer.⁶²

183. The availability of a computer may allow audio-visual media to be used, as well as words. This might be an effective way of testing certain cognitive skills. On the whole, giving respondents a variety of tasks is likely to maintain their interest for longer than repetition of a particular style of data collection. However, this might be hindered by lack of features in certain software packages and, if a lowest common denominator approach is required, this means using computers in text mode.

184. It is worth noting in this context it is not necessary for all data to be collected on the computer. One may continue to use printed materials, either as text prompts ('Show Cards') or with graphics, or as a self-completion booklet. On the whole, however, it will be more cost-effective and will generate better-quality data if the respondent can use the computer.⁶³

⁶² Usually, the pre-coded answers will be displayed under the questions. Respondents may just look at the answers and assume they can infer what the question is. Some CAPI software allows all keyboard activity to be logged. This means one can calculate how long a respondent spent in answering each question, and also if they changed their answer.

⁶³ Self-completion on the computer allows more complex decisions to be made about the sequence of questions to be answered, as well as enforcing adherence to answering every question in the right order. Ticking boxes on a paper sheet allows none of these, with possible adverse effects on data quality.

APPENDIX E. AN OUTLINE OF INITIAL COST ESTIMATES AND RELATED ASSUMPTIONS

185. This appendix outlines the basis on which a ‘model PIAAC costing’ has been generated for the UK at 2004-05 costing rates. These costs are expressed in Euro.

186. Four alternative survey designs are costed:

- A cross-section of the workforce, across the entire age spectrum
- A survey focusing on young people, defined here as those aged 18 to 35.
- A survey focusing on older workers, 45 to retirement age (however, the costings for this option are here assumed to be identical to those of the option focusing on young people, although in practice the costings are likely to vary between the two age groups).
- A survey of a cross-section of ages, with over-sampling of a specific age group

The costing assumptions are explained in four sections:

- i) Those costs that are common to all strategy options.
- ii) Those costs that are specific to the ‘entire age spectrum’ option.
- iii) Those costs that are specific to the younger/youth age group (18-35) and the older workers age group.
- iv) Those aspects of costs that are specific to a survey of a cross-section of ages, with over-sampling of a specific age group.

E1. Assumptions that are common to all options

- Overall achieved sample size: 4,000 interviews, conducted in-home (however, the tables indicate the approximate costs associated with samples of 2,500 and 7,500, treating many elements as fixed even though, in practice, they would vary to some extent with the scale of the survey).
- Length of interview: median 60 minutes, 90th percentile approx. 80 min.
- Mode of interview: CAPI, with about half self-completion by the respondent.
- No incentive for respondents.
- Sample clustered in about 200 Primary Sampling Units (PSUs), with a uniform issued sample per PSU.
- Number of interviewers about 200 (excluding supervisors), and each interviewer to attend one day of briefing.
- Response rate, approx. 65 per cent on residential addresses where someone is identified who is eligible for the study.

- Letter sent in advance by postal service to all addresses.
- Pilot survey of 100 issued and 40 achieved as ‘dress rehearsal’ exercise.
- Questionnaire incorporates key checks on consistency of data.
- SOC/SIC and other specialised coding (approx. 8-10 minutes of coding/editing per interview) conducted after data collection.
- Output is an anonymised, weighted, edited, labelled data file in a compatible format on CD or DVD.
- Research staff time for managing data collection estimated at 200 person-days (assuming a single-stage fieldwork operation).
- Research staff time includes technical report (as PDF file) to document sample design, timetable of project, weights, etc, and to provide questionnaire documentation.
- No interpretative report included in estimate of research staff time.

E2. Assumptions specific to entire age spectrum option

- 6,000 eligible addresses identified (with a resident member aged 18 to 65 and currently or recently in paid work), to achieve 65% response rate and allowing for one interview per address (with variants for 2,500 and 7,000 interviews).
- Issued sample⁶⁴ has approximately 45% of residential addresses ineligible, thus approximately 11,400 addresses selected (57 per PSU), distributed in proportion to population.
- Two achieved interviews per 6-hour day (due to screening and unproductive addresses, this estimate is exclusive of quality control⁶⁵).

E3. Assumptions specific to the option of focusing on youth (aged 18-35) and the option focusing on older adults

- The key assumption relates to the proportion of residential addresses at which there are one or more people aged 18 to 35 (or 45 to retirement age) and who are in paid work or have worked within a recent period; for Britain, this is assumed to be approximately 17 per cent of addresses.

⁶⁴ It is not an essential feature of this project that all the issued sample be handled at one time; it would be possible to reduce the number of interviewers to be briefed and supervised by dividing the sample into 3, 4 or more parts.

⁶⁵ In Britain, one of the standard fieldwork quality control measures is to re-contact a proportion of the addresses, to check how the interview was conducted; another is to observe the work of the interviewer while he/she is conducting interviews; the work of all interviewers is checked every 6-12 months, so these checks would not apply to all the interviewers working on this project. Organisations in other countries may implement different quality control measures and should allow in costing for the standard measures used in their country.

- 6,000 eligible addresses identified (with a resident member aged 18 to 65 and currently or recently in paid work), to achieve 65% response rate and allowing for one interview per address (with variants for 2,500 and 7,000 interviews).
- Issued sample of 44,000⁶⁶ addresses, in 8,800 sets of 5 addresses; for each set of 5, at least one address to be contacted⁶⁷ in order to ‘screen’ all 5 addresses for people who may be in the eligible age range, expecting to yield 6,000 addresses identified as eligible.
- It is permitted to conduct five interviews within one set of addresses in the event that all five addresses have someone eligible; however, there is a maximum of one interview per address.
- The rate per day is 1.5 achieved interviews, allowing for the time required for ‘screening’ at the sets of 5 adjacent addresses (as well as interviewing).
- No initial letter sent by post to the sampled addresses (as we expect that around 5 in 6 will be screened-out and need not be aware that they had been selected).

E4. Cost estimate for survey of 4,000 adults aged 18 to 65

The following table provides indicative costs for three different sample sizes at costing rates current in the United Kingdom in 2004-05. The different sample sizes help to clarify which of the costs are effectively ‘fixed’ and which vary according to the size of the sample. In practice, the costs shown as fixed would vary to some extent according to the sample size.

Key assumptions applying to all three sample sizes are:

- The supplier’s responsibility is to manage the fieldwork operation and supply an edited and coded data file (no analysis or reporting is included).
- The questionnaire has been specified fully, but requires a pilot survey to check operational arrangements in advance of the main survey.
- The incidence of eligible people is expected to be around 55 per cent of residential addresses.
- Eligible people are to be identified by contacting each address.
- The length of interview is expected to be 60 minutes.

⁶⁶ This initial sample allows an excess of almost 25 per cent of sample, to recognise that interviewers will be misinformed by some people; this allows a margin of error for the key estimate of someone 18-35 being found at one in six addresses.

⁶⁷ Where the person first contacted is unable to say whether anyone may live in a selected address, then the interviewer has to seek another person to ask, including calling at any address about which there remains uncertainty; only ‘responsible adults’ may be asked about the current residents of addresses.

Research activity	Achieved sample size		
	2500	4000	7500
Research staff	189000	189000	189000
Computing staff	26100	26100	26100
Fieldwork - Development	4800	4800	4800
Fieldwork - piloting	6750	6750	6750
Fieldwork - Main survey	517500	780000	1245000
Fieldwork - Other costs	54750	65700	82500
Office clerical and handling	63000	90000	120000
Expenses			
Printing	3000	3300	3900
Sampling	6000	7500	9750
Postage, despatch	300	300	450
Travel, hotel	2250	2250	2700
Other costs			
Overall total (Euro)	873450	1175700	1690950
Cost 'per interview' (Euro)	349	294	225
Rate per day	2	2	2

Note: Costs based on costing rates in 2004-05

E5. Cost estimate for survey of 4,000 adults aged 18 to 35 (with identical costs assumed for the age group 45-65)

187. The following table provide indicative costs for three different sample sizes at costing rates current in the United Kingdom in 2004-05. The different sample sizes help to clarify which of the costs are effectively 'fixed' and which vary according to the size of the sample.

Key assumptions applying to all three sample sizes are:

- The supplier's responsibility is to manage the fieldwork operation and supply an edited and coded data file (no analysis or reporting is included).
- The questionnaire has been specified fully, but requires a pilot survey to check operational arrangements in advance of the main survey.
- The incidence of eligible people is expected to be around 17 per cent of residential addresses.
- Eligible people are to be identified by 'screening' sets of five adjacent addresses.
- The length of interview is expected to be 60 minutes.

188. It is expected that costs would be very similar if the age group was 45-65.

Research activity	Achieved sample size		
	2500	4000	7500
Research staff	189000	189000	189000
Computing staff	26100	26100	26100
Fieldwork - Development	4800	4800	4800
Fieldwork - piloting	6750	6750	6750
Fieldwork - Main survey	682500	1003500	1665000
Fieldwork - Other costs	54750	65700	82500
Office clerical and handling	63000	90000	120000
Expenses			
Printing	3000	3300	3900
Sampling	6750	9000	13500
Postage, despatch	300	300	450
Travel, hotel	2250	2250	2700
Other costs			
Overall total (Euro)	1039200	1400700	2114700
Cost 'per interview' (Euro)	416	350	282
Rate per day	1.5	1.5	1.5

Note: Costs based on costing rates in 2004-05

E6. Cost estimate for survey of 4,000 adults aged 18 to 65, with 100% extra 18 to 35

189. The following table provides indicative costs for three different sample sizes at costing rates current in the United Kingdom in 2004-05. The different sample sizes help to clarify which of the costs are effectively 'fixed' and which vary according to the size of the sample.

190. Key assumptions applying to all three sample sizes are:

- The supplier's responsibility is to manage the fieldwork operation and supply an edited and coded data file (no analysis or reporting is included).
- The questionnaire has been specified fully, but requires a pilot survey to check operational arrangements in advance of the main survey.
- For the middle sample size, 7000 in total, the process would start with 8,200 addresses, which would be visited by interviewers. We assume that 55% of these addresses would have someone aged 18-65 and in work (i.e. 4,500 eligible individuals) at which a 65% response rate would yield around 3,000 interviews.

- Along with each of the 8,200 starting addresses, interviewers would be given the two preceding and the two following addresses from the Postcode Address File. Irrespective of whether they found an eligible person at the middle ('core') address, they would ask about the residents of the two addresses on either side, to see whether any of these four addresses contained anyone aged 18-35. Where they were told 'No', they would remove the address from the sample. If the answer was 'Yes' or 'Don't know', they would visit the address and verify the information. This gives an additional 32,800 addresses to be checked. If the incidence of people aged 18-35 was 20%, then this translates as 6,560 eligible people. Assuming a response rate of 65% among these 6,560 eligible people would give the required 4,000 interviews (in fact, it would provide another 264, but a margin for uncertainty in such approximate numbers is desirable). The cost of the additional screening work is fairly modest, as the aim would be to identify the neighbouring addresses in the course of dealing with the core sample. It is assumed that the rate per day would be 1.6 60-minute interviews (as compared with 1.5 for screening only, 1.7 for a smaller screening operation and 2.0 per day for 18-65s only).

Research activity	Achieved sample size		
Core sample, adults aged 18-65	1500	3000	4500
Boost sample, adults aged 18-35	2000	4000	6000
Total number of interviews	3500	7000	10500
Research staff	189000	189000	189000
Computing staff	26100	26100	26100
	0	0	0
Fieldwork - Development	4800	4800	4800
Fieldwork - piloting	9000	9000	9000
Fieldwork - Main survey	825000	1650000	2475000
Fieldwork - Other costs	60150	120000	181500
Office clerical and handling	54750	108000	159000
Expenses	0	0	0
Printing	3600	4950	7200
Sampling	6750	9000	13500
Postage, despatch	750	1200	2100
Travel, hotel	4500	6750	9000
Other costs	0	0	0
	0	0	0
Overall total (Euro)	1184400	2128800	3076200
Cost 'per interview' (Euro)	338.4	304.1	292.9
Rate per day	1.6	1.6	1.6

Note: Costs based on costing rates in 2004-05