BEYOND PISA 2015: A LONGER-TERM STRATEGY OF PISA

Introduction

1. The paper sets out the purpose and principles of PISA and six objectives for the longer-term direction of PISA beyond PISA 2015. The paper is based on discussions made at the PISA Governing Board.

The purpose and principles of PISA

2. Launched in 1997, PISA responds to the need for cross-national comparisons on student performance. It aims to provide reliable information on how well education systems prepare students for further study, careers and life. PISA also provides a basis for international collaboration in order to define and implement effective educational policies.

3. When considering the long-term direction of PISA it is imperative to take a look at the following principles that have been the driving force behind the development of PISA:

a) PISA is a *system-level assessment.* It represents a commitment by governments to monitor the outcomes of education systems. The assessment's main outputs provide internationally comparable evidence on the quality, equity and efficiency of school systems at the system level.

b) PISA is *policy-oriented.* It focuses on providing data and analysis that can help guide decisions on education policy. By linking data on students' learning outcomes with data on key factors that shape learning in and out of school, PISA highlights differences in performance patterns and identifies features common to high-performing students, schools and education systems.

c) PISA is carried out *regularly* to enable countries to monitor their progress in meeting key learning objectives. The basic survey design has remained constant to allow for comparability from one PISA assessment to the next and thus to allow countries to relate policy changes to improvements in education outcomes.

d) PISA assesses both *subject matter content knowledge*, on the one hand, and *the capacity of individuals to apply that knowledge creatively*, including in unfamiliar contexts, on the other. Until now PISA has assessed students in three main domains in each round: reading, mathematics and science. These domains cover some of the fundamental skills needed to continue on with higher education or succeed in the labour market. In addition, one innovative assessment domain is typically included in each cycle.

e) PISA *focuses on knowledge and skills towards the end of compulsory schooling*. It implements and analyses internationally standardised assessments of student learning outcomes up to the end of compulsory education. In most countries the end of compulsory education is around the age of 15 where students are supposed to have mastered the basic skills and knowledge to continue on to higher education or in the workforce.

f) PISA is designed to provide comparable data *across a wide range of countries*. Considerable efforts are devoted to achieving cultural and linguistic breadth and balance in assessment materials. Stringent quality-assurance mechanisms are applied in the test design, translation, sampling and data collection. An age-based rather than a grade-based target population is used to ensure valid international comparisons of educational performance.

g) PISA is *a collaborative effort.* Decisions about the scope and nature of the PISA assessments and the background information collected are undertaken by leading experts in participating countries. Governments oversee these decisions based on shared, policy-driven interests. New initiatives are considered in terms of their consistency with the longer term strategy that applies at a given time.

4. The innovative assessment domain for each cycle, as mentioned in d) above, can be either a cognitive assessment, such as the assessment of problem-solving competencies in 2003, 2012 and 2015, or an assessment of non-cognitive dispositions such as the self-assessment of learning strategies in 2000 or the assessment of student attitudes towards science in 2006. The innovative domain can also cover a combination of cognitive and non-cognitive elements. The domains developed in previous cycles have been either cross-curricular or have had a close relationship with the major domain of a given cycle. The international costs of development and implementation of the innovative domain are part of the core PISA budget. The domain for each cycle is chosen in terms of:

- The ability to provide insight in a new area, as has been the case with the evolution of problem solving from paper-based in 2003 to computer-based individual problem solving in 2012 to computer-based collaborative problem solving in 2015.

- Consistency with the principles of PISA, as set out above, and the objectives for further development of the study, as set out in the remainder of this document.

- Relevance to education policy, including the ability to provide information on how well students are prepared for full participation in society.

- Feasibility of the assessment, including the practical constraints on student testing time and financial resources.

5. This strategy also takes account of the outcomes of the discussion at the informal meeting of OECD Ministers of Education in October 2013 regarding the future direction of PISA. Ministers at the meeting acknowledged the unique nature of PISA as a tool for policy making, welcomed the progress that had been made in driving PISA forward and recognised a number of areas for improvement which are in line with the objectives in the present longer term strategy.

The economic and societal developments shaping the future demand for skills

6. When considering the long-term directions of PISA, it is also relevant to look at the societal and economic transformations that are reshaping the working and social conditions of young adults. This section gives a brief overview of these transformations and details the likely impact they will have on the demand for skills and competencies.

7. Rapid globalisation and modernisation are posing new and demanding challenges to individuals and societies alike. Increasingly diverse and interconnected populations, rapid technological change in the workplace and in everyday life, and the instantaneous availability of vast amounts of information are just a few of the factors contributing to these new demands. In this globalised world, people compete for jobs not just locally but internationally. In this integrated worldwide labour market, there are many occupations in which highly-paid workers in wealthier countries are competing directly with people with much the same skills in lower-wage countries. The same is true for people with low skills. Competition among countries now revolves around the quality of their human capital.

8. But these developments do not affect all workers equally. Job automation is proceeding even faster than the integration of the job market. If the work is routine, it is increasingly likely to be automated,

although some jobs will always be done by human beings. The effect of automation, and more generally of the progress of technological change, is to reduce the demand for people who are only capable of doing routine work, and to increase the demand for people who are capable of doing knowledge-based or manual work that cannot be expressed in the forms of algorithms. High-wage countries will find that they can only maintain their relative wage levels if they can develop a high proportion of highly skilled workers and keep them in their work force. Increasingly, such work will require very high skills levels and will demand increasing levels of creativity and innovation.

9. This is not a description of one possible future, but of the economic dynamics that are now in play. In the high-wage countries of the OECD, demand for highly skilled people is increasing faster than supply (which OECD indicators show in rising wage premiums for highly-skilled individuals); and demand for low-skilled workers is decreasing faster than supply (which OECD indicators reveal in growing unemployment rates or declining wages for low-skilled individuals). Jobs are moving rapidly to countries that can provide the skills needed for any particular operation at the best rates. And the rate of automation of jobs is steadily increasing in both high- and low-wage countries.

10. The transformation of labour markets, from a high reliance on routine and manual work to knowledge-based intensive work is continuing and, aided by new technologies, it is profoundly reshaping the nature of workplaces. Individuals are no longer expected to passively consume information coming from well-defined sources and to use the knowledge they accumulated in the ways they thought would be useful when they were developing it. Information is now produced by a multitude of conflicting sources, and knowledge needs to be transformed and applied to novel situations. The knowledge workers of today are required to have deep knowledge, but the knowledge workers of tomorrow will need deep and wide knowledge: knowledge that can be moulded and shaped to fit a transforming world. The need for deep and wide knowledge means that education systems will need to help students adapt to new situations by giving them a mindset that is ready to absorb and filter new information and is able to combine new information with acquired knowledge in new, innovative ways. More than ever education systems need to help students learn how to learn: only if students have the capacity, the motivation and enthusiasm to be life-long learners will they be able to remain active and productive citizens throughout their lives and reap the full benefits that life offers.

11. The task of equipping students with the ability to be life-long learners is compounded by demographic trends. Declining fertility and increasing life-expectancy worldwide mean that populations are aging. Economic growth and stability will depend on the ability of workers to remain in the labour force and to continue having high levels of productivity for longer. Demographic trends will also shift population pyramids towards a smaller base of young, active workers. It will therefore be increasingly important for education systems to tackle barriers that prevent some students from achieving their full potential. For example, socio-economically disadvantaged boys too often drop out of formal education with few skills and, even more worryingly, little willingness and motivation to develop them in the future. Education systems have also so far been unable to make sure that the large numbers of girls who have the ability to excel in mathematics are willing and able to develop their potential to go on and fill occupations in rapidly developing STEM industries. Unless education systems develop and capitalise on the talent of each and every student, demographic changes mean that countries as a whole will likely experience shortage of skills in the future. Never before have equity of educational opportunities and economic efficiency been so closely intertwined.

12. In this context, governments need to create education systems that are globally competitive in quality; that provide people from all classes a fair chance to get the right kind of education to succeed; and to achieve all this at a price that the nation can afford. Such education will need to build the very high skills levels required to solve complex problems never seen before, to be creative, to synthesise material from a wide variety of sources, to see patterns in the information that computers cannot see, to work with

others in productive ways, and to be able to both lead and be a good team member when necessary. This is what is required in today's "flat" world – where all work that cannot be digitised, automated and outsourced can be done by the most effective and competitive individuals, enterprises or countries, regardless of their location. The implication is that the yardstick for educational success is no longer simply improvement against national standards, but against the best-performing education systems worldwide.

13. It is also worth taking into account the reflections and developments that have come out of the global financial and economic crisis. New sources of growth are necessary to put economies at a strong, inclusive and sustainable growth path to support the wellbeing of populations. Schooling, in general terms, and academic achievement, in particular, are means to achieve the wellbeing of citizens and not only a goal in itself. Competencies of individuals are important not only because of their contribution to economic growth, employment and innovation, but also because of their importance in other domains. Outside the domains of economics and work, the competencies of individuals contribute to increasing participation in democratic institutions, and to strengthening social cohesion, human rights and autonomy as counterweights to increasing global inequality of opportunities and increasing individual marginalisation.

14. Future developments of PISA may benefit from following the outcomes of the New Approaches to Economic Challenges (NAEC) reflection process launched in 2012. NAEC is aimed at continuously improving analytical frameworks and policy advice of OECD's work, given the interlinked policy challenges related to a hesitant recovery from the crisis. Also the framework developed by Rychen (2003) as part of the *Definition and Selection of Competencies* programme is relevant in this context. It covers competencies in three broad categories: interaction in socially heterogeneous groups, the capabilities of acting autonomously and using tools interactively. Trier (2003) and Binkley, Sternberg, Jones and Nohara (1999) provide other relevant frameworks for the measurement of key competencies.

Objectives for the longer-term development of PISA

15. With the context set out in the previous sections in mind, the remainder of the paper seeks to set out the directions for the longer-term development of PISA. It proposes six objectives, each of which touches upon a different aspect of the programme.

16. There is general agreement in the PISA Governing Board to maintain the basic principles of PISA and build on the survey's success through improvements within its existing structure. The following six objectives are therefore conditioned by the continuation of the principles outlined in the first section of this document.

17. The six objectives for the longer-term development of PISA are:

> ***Objective 1***
> **To continue exploiting the advantages of computer-based testing, including increasing the use of adaptive testing, allowing to progressively diminishing the distinction between major and minor domains, and facilitating the better measurement of non-cognitive outcomes**

18. Computer-based testing has been gradually introduced to PISA since the first optional electronic module in 2006. The transition to computer-based testing makes it possible to broaden the scope of the skills and competencies measured, to reflect the integration of new technologies in education and in the society as a whole, and to enhance the measurement quality and efficiency. PISA 2015 marks a significant step in this development with the majority of countries moving to a fully computer-based assessment platform.

19. Looking beyond PISA 2015, PISA shall continue to exploit the advantages of computer-based testing and other technological and methodological innovations. These developments could include further use of computer-based adaptive testing and new and innovative item formats in all assessment domains. Technological innovations may also be used to expand existing assessment domains and assess new constructs that are difficult to measure through traditional paper-based assessments.

Adaptive testing and a progressive diminishing of the distinction between major and minor domains

20. Since the first cycle, PISA has focused on three core domains with one domain identified as the major domain in each cycle and the other two treated as minor domains, plus one or more optional domains. The amount of time allotted for the measurement of each domain is determined by this distinction. As less testing time is devoted to the minor domains, results are currently not available at the subdomain level.

21. This disadvantage may be addressed by the introduction of computer-based adaptive testing in which the computer chooses the items (or clusters of items) that are given to each student so they fit the students' ability level, based on the student's performance on earlier items. Such an approach can increase the efficiency of the assessment and, potentially, be used for reporting on subdomain level for reading, mathematics and science in each cycle.

22. The proposed survey design for the main study of PISA 2015 includes some adaptivity at the cluster level, although most likely not to an extent that will allow for reporting on the subdomain level for all subjects. For future PISA cycles, adaptive testing may be further introduced in order to fully exploit the advantages of the methodology. PISA was not from its inception designed to be an adaptive assessment and so considerable development will be needed if the assessment is to fully exploit the benefits of the methodology while ensuring trend measurement and continued international comparability. With a progressive introduction of adaptive testing, PISA may also benefit from the developments of the methodology in other assessments and the educational research that may be undertaken to evaluate such developments.

23. It is worth noting that the potential gains in measurement efficiency resulting from adaptive testing may be used for different purposes. Along with the technical developments of the methodology, the strategic objectives of adaptive testing will need to be considered in order to prioritise the use of such gains. Improving the measurement of minor domains to the level of major domains is one possible objective, while other possibilities include reducing the assessment time or enhancing the measurement for the highest and lowest performing students.

24. Computer-based testing will also make it possible to address another disadvantage of minor domains. In the design that has been used in the paper-based assessment, the number of items for the minor domains has been considerably smaller than for the major domains. With a computer-based design, the items can be more easily distributed between the students allowing for an increase in the number of items in each domain. In PISA 2015, the number of items in the minor domains, all of which are link items, will ensure a construct representation more comparable to that of a major domain. This reduces the potential bias in the measurement of trend, without increasing the sample size, the assessment time or weakening the major domain. Looking beyond PISA 2015, a larger number of items could be carried forward from major to minor domains, depending on whether the current structure of major/minor domains continues.

Better measurement of non-cognitive outcomes

25. Non-cognitive and non-subject-specific outcomes have been an important element of PISA since the programme's beginning. Computer-based testing provides at least three ways of improving the measurement of these outcomes First, new questionnaire item formats may provide more precise and

relevant measurements as well as a more engaging experience for students as they fill in the questionnaires. New item formats could include scenario-based questions, video clips of a specific situation, or interactive visual representations.

26. Secondly, the validity and coverage of the cognitive questionnaires can be enhanced by collecting and analysing process data through the computer-based questionnaires. The major goal of using process data from the questionnaires is to consider response sequences and response times to detect aberrant response and speediness, to derive indicators for additional constructs, and to evaluate and assure the quality of the questionnaire design. Apart from the non-cognitive outcomes, process data from the cognitive assessment may also be used to further explain and describe student performance, for example to describe the different task solution strategies that students use to solve a mathematical problem.

27. Thirdly, computer-based assessment may be used to facilitate a closer integration of the assessment of non-subject-specific competencies with the main cognitive test. Non-subject specific competencies not only play a central role shaping students' ability to achieve at high levels in these subjects, they also represent valuable attributes that will enable students to succeed in the work place and lead full lives, providing them with the ability to successfully deal with challenges and to make the most of available opportunities. The clear separation in the PISA design between the questionnaire and the test booklets have so far limited the possibility for PISA to examine non-subject specific skills in a systematic way and derive information on student dispositions and behaviours in a way that is not based on students' self-reports. Computer delivery and methodological innovations could create opportunities for the development of a more integrated PISA assessment aimed at assessing dispositions rather than asking students to report their attitudes and behaviours. For example while the 2012 PISA student questionnaire contains questions on perseverance, future cycles could measure perseverance using specific test items rather than rely on students' self-reports. It would also be worth exploring if process data from the cognitive test can provide information on non-cognitive outcomes.

The role of the internet in PISA

28. Lastly, the rapid expansion of internet access will make it relevant to explore and consider the internet's role in PISA. Some countries are already experimenting with providing students with access to internet resources in national tests and examinations. Internet access may increase the validity of an assessment of students' abilities to solve tasks in real-life settings, but could introduce risks of cheating or plagiarism for certain item types.

29. The rapid growth of global internet access may also present an opportunity for online delivery of the assessment, with or without giving students web access during the assessment. In PISA 2015 the school and teacher questionnaires will be delivered via the Internet and assessment materials and results data will to some extent be transferred via the Internet between schools, national centres and central servers. However, the test itself is not delivered online in real time. This is an option that it may be worth exploring for future PISA cycles, as internet access across schools is likely to become more widespread.

Objective 2

To aim for greater synergies between PISA and OECD and other international surveys and data collections on education in order to provide more coherent information about learning outcomes as students move through the education system

30. As the largest international education survey, PISA is in a key position to pursue synergies with other international surveys and data collections on education. The aim is to enhance PISA's analytical power and policy relevance by providing more coherent and reliable information about learning outcomes throughout the education system.

31. An example is the greater coherence between PISA and the Teaching and Learning International Survey (TALIS), which is currently being explored in terms of the design of the measurement instruments and the operational co-ordination of the surveys. TALIS was established to provide a broader perspective from teachers on issues around the teaching profession. Synergies between PISA and TALIS can enhance data quality and yield more and better policy insights from both studies. The PGB has agreed to seek greater coherence between the measurement of constructs common to TALIS and PISA and to establish a common expert group for questionnaire development. As a result, a number of items for the optional teacher questionnaire in PISA 2015 have been taken from the TALIS 2013 questionnaire when the purpose of the items was the measurement of the same constructs for both programmes (e.g. for measuring teacher background information and teacher self-efficacy). Further potential synergies are being explored and discussed by the governing boards of both studies. Increased synergies can be achieved by aligning one or more specific elements of the programmes, such as the survey cycles, the samples or the analysis plans. Such alignments would in some cases enable triangulation of data from different sources at the school level, *i.e.* analyses using both PISA and TALIS data.

32. Another example is the potential synergies between PISA and the Programme for the International Assessment of Adult Competencies (PIAAC). The two programmes are complementary. While PISA measures the learning outcomes at age 15, the last point at which schooling is still largely universal, PIAAC begins at age 16 and extends to the age of 65 years. While PISA is looking backwards to establish how effectively school systems establish the foundations for success in life, PIAAC is looking forward to how initial skills feed into further learning and important economic, employment and social outcomes. The combination of the two studies provides unique opportunities to analyse how learning outcomes at school shape the way students go through further education and on to the labour market. In one country, a cohort of 1500 students who participated in PISA 2000 were included in the PIAAC sample in 2011-12 in order to analyse the relationship between school outcomes and adult skills, educational career choices and the employment situation at the of age of 27. Not all countries have the opportunity to make such longitudinal studies, but links can also be made without following a specific cohort. At the instrument level, the two assessments could be closer linked by developing common items or by conducting an equating study to analyse the correlations between the scales, and if feasible, to create a statistical link between the PISA and PIAAC scales.

33. In-depth comparisons between PISA and other surveys may also be initiated in order to provide policy makers and education practitioners with more coherent information regarding the results obtained through different surveys. One such in-depth comparison, published in 2010, analyses and explains the observed differences and similarities between the results in students' mathematics performance in PISA 2003 and TIMSS 2003 (Wu 2010). The study also focuses on the complementary findings from the two surveys and the lessons which can be learned from a cross-comparison in order to improve each survey.

34. Table 1 provides an overview of the four large OECD-based education studies (PISA, PIAAC, TALIS and AHELO) as well as the three largest international IEA studies in terms of participating countries (PIRLS, TIMSS and ICCS). As the table shows, most age groups are covered in some form or other by large international educational studies. Early childhood has not been covered by international studies up to now, but different studies are currently being developed. For instance, the PGB has been following progress with the development of iPIPS, an international study on learning outcomes in early childhood education co-ordinated by Durham University. In this context, the PGB has shown interest in developing links between PISA and early years assessments and has noted that such links would most likely be based on synthetic age groups rather than longitudinal cohorts. The Strategic Development Group will continue to assess the utility of iPIPS and other instruments in the context of PISA.

Table 1. Major international education surveys

Study	Age	Subject areas	Sources of context information*	Frequency	Global coverage
OECD PISA	15	- Reading - Mathematics - Science - Collaborative problem-solving (2015) - Problem-solving (2012) - Financial literacy	- Students - Parents (optional) - Teachers (optional) - School principals	Every three years since 2000	OECD countries: 34 Non-OECD participants: 40 (PISA 2009)
OECD PIAAC	16-65	- Literacy - Numeracy - Reading components - Problem-solving in technology-rich environments	- The individuals who are assessed	Frequency to be decided**	OECD countries: 24 Non-OECD participants: 2 (PIAAC 2011)
OECD TALIS	Teachers of lower secondary education***	(Focuses on the learning environment and working conditions of teachers)	- Teachers - School principals	Five years between first two cycles	OECD countries: 16 Non-OECD participants: 7 (TALIS 2008)
OECD AHELO	University students at the end of the B.A. programme.	- Generic skills common to all university students (such as critical thinking) - Skills specific to economics and engineering	- Students - Faculties - Institutions	A feasibility study was carried out in 2012	Institutions from 17 countries participated in the feasibility study in 2012
IEA PIRLS	4th grade	- Reading	- Students - Parents - Teachers - School principals	Every five years since 2001	OECD countries: 24 Non-OECD participants: 26 (PIRLS 2011)
IEA TIMSS	4th and 8th grade students	- Mathematics - Science	- Students - Teachers - School principals	Every four years since 1995	OECD countries: 26 Non-OECD participants: 36 (TIMSS 2011)
IEA ICCS	8th grade****	- Knowledge and understanding of civics and citizenship	- Students - Teachers - School principals	First round in 2009, second round in 2016	OECD countries: 23 Non-OECD participants: 15

* Most studies also collect system-level information
****** The option of a 10 year cycle with interim studies after 5 years is being explored
*** With an option to expand the survey to elementary and upper secondary schools in TALIS 2013
******** Provided that the mean age at the time of testing was at least 13.5 years

Objective 3

To continue to seek methodological and analytical means to strengthen the policy relevance and analytical power of PISA, including establishing best practice for linking PISA with national assessments

35. Several options for strengthening the policy relevance and analytical power of PISA have been discussed in the development of PISA over the last years and shall continue to be developed for future cycles. These include improvements of the context questionnaires, the establishment of best practice for linking PISA with national assessments and the options for strengthening the ability to draw causal inferences from the data.

36. The validity of the questionnaires can be improved through the use of new, innovative item types, which were introduced in PISA 2012 and further developed with the transition to computer-based delivery in PISA 2015. One approach, which has already been introduced, is to replace Likert-type questions (typically measured on a scale from "strongly disagree" to "strongly agree") with alternative item types such as forced choice questions or situational judgement tests. These item types can improve the cross-cultural comparability of existing measures and reduce the effects of response styles. If the move towards computer-based delivery is continued in future cycles, this could open up more opportunities for innovations that could further improve the validity of the questionnaire data. As mentioned in Objective 1, more precise and relevant measures may be obtained by including scenario-based questions and interactive visual representations. One advantage of such item types would be to reduce the text load and thus potentially decrease the impact of reading competencies on student responses. Further use of computer delivery will also make it possible to improve the data quality of questions asking for time use in and out of school. More reliable responses could be obtained for example by displaying an interactive weekly schedule or timetable in which students can tick days and hours for different lessons or out of school activities. 37.PISA's analytical power may also be improved by widening the sources of information used in the analyses and exploring alternative methods for collecting the contextual data. Some background information about students might be obtained by using existing national data available through administrative archives. Another source that could be further explored is the use of system-level data. DIPF has proposed that system-level data be used as a measure of attitudinal differences at the country-level to address the paradoxical finding that while attitudinal outcomes based on the student questionnaire are positively correlated with achievement at the student level, the correlations are in some cases negative at the country level. With regards to science, system-level indicators could include the number of science-related TV and radio programmes, or the proportion of adults working in science- and technology-related professions.

38. The optional teacher questionnaire introduced in PISA 2015 is also an additional source of information which can improve PISA's ability to explain variation in educational outcomes. The questionnaire responds to a growing interest in teacher and teaching-related policies in participating countries. While in previous cycles school principals have provided some information about teachers, the teacher questionnaire will enable PISA to have information from teachers themselves, which will yield more valid, more detailed, and more sophisticated school-level information and significantly enhance the analytical power of PISA in many policy areas. The questionnaire covers such topics as teachers' perspectives on their working conditions, characteristics of the teaching workforce, information on subject-specific topics and school-level information about context and the learning environment.

39. It has been suggested that these and other improvements of the questionnaires for future cycles could be investigated through a research seminar with experts already involved in PISA as well as other experts. Such a seminar could focus on improving and validating measures of social background; adding system-level data; cross-national validity; and validating and improving trend measures.

40. In relation to these issues it may also be relevant to consider increasing the amount of contextual information gained through the background questionnaires, in particular the student questionnaire. Areas that could be further explored include the psychological and social relation outcomes of schooling and the satisfaction of students and families with the school and the education system overall. One possible option to increase the amount of student data collected is to include more questions in the questionnaire and similarly decrease the amount of time spent on the cognitive test. However, this option would only be relevant if the robustness of the cognitive assessment can be retained and continue to be developed. It is possible that adaptive testing, if it can be successfully implemented in PISA, may make it possible to decrease the length of the test without decreasing its robustness, thereby allowing for an increase in the length of the questionnaire. A second option is to use a rotated design for the student questionnaire. The results of the 2012 rotated questionnaire design could be analysed further to ascertain whether a similar

design would prove suitable in future cycles. A third option may be to select a number of topics in the questionnaires that do not need to be measured every three years, but could be monitored over a longer time scale in order to free up space for other areas to be explored in greater depth.

41. Another option to improve PISA's analytical power is through strengthening the survey's ability to draw causal inferences. Different analytical techniques can improve the effect estimation when analysing what factors are related to educational outcomes. Such techniques include panel models that use information of changes over time, effect analyses that use an instrumental variables approach and regression discontinuity models, which can be used for example to enhance the estimation of the effect of one year of schooling.

42. One such method is the establishment of school panels, *i.e.* assessing the same schools in two or more cycles. Some countries are already working with such panels. The method can be used to identify changes in the instructional context or student composition at the school-level and relate them to changes in school performance over time. The longitudinal information can sometimes provide information about causal inferences that is more powerful than what can be obtained in a cross-sectional study in which both the context variables (*e.g.* the school climate) and the outcome variables (student achievement and attitudes) are measured at the same time. Another cost-effective way to obtain longitudinal data may be to assess the same schools or even the same students in the field trial and the main study of PISA. This approach would provide information of changes in cognitive and non-cognitive outcomes over one year and information of how these changes can be predicted based on school and student level predictors. However, the reassessed students would not be eligible for the standard PISA sampling, and establishing such a link would therefore only be an option for countries that volunteer to increase their samples.

43. Associated with causal inferences, methods to combine policy relevance with methodological rigour in the PISA reports shall continue to be elaborated. For example, differences in the strengths of inferences drawn from the PISA data might be more effectively highlighted in the PISA reports by classifying the robustness of inferences drawn from PISA data and presenting that classification in the reports.

44. Finally, the policy value of PISA can be improved at the country level by linking PISA with national assessments and thereby enabling comparisons between the learning results achieved through the national assessments and the international standards established by PISA. A survey among participating countries has shown that at least five countries have implemented such links. Different methods can be applied, including statistical linking through common items or the application of a common linking sample to PISA and the national test. Linking PISA with national assessments is also a possible way to obtain longitudinal information, since national assessments often have a longitudinal aspect. A paper on practices of linking PISA with national standards and assessments and another paper on the use of PISA to internationally benchmark state performance standards have been prepared over the last two years [doc. ref. EDU/PISA/GB(2011)/17]. This work could be continued in order to add further detailed case studies and produce a detailed methodology paper giving best practice advice to countries.

Objective 4
Broaden the global coverage of PISA

45. Broadening the global coverage of PISA will increase the survey's capacity to provide international comparisons and insights from a variety of countries and economies. Most importantly, it will help countries that join the programme to establish themselves on an improvement trajectory to achieve targets referenced to common international goals. Increased participation will also facilitate the tracking of international education targets across new-coming countries.

46. However, it is important that efforts to broaden the coverage of PISA be done without compromising PISA's capacity to measure what it measures today. PISA should continue to develop according to the principles and purposes of the programme, as set out in the beginning of this document. Any improvement of the main PISA instruments in order to increase the number of participating countries will therefore be done with due consideration to maintaining the programme's relevance to OECD countries and other existing participants.

47. The PISA Global Relations Strategy sets out the processes, procedures and criteria for engaging with partner countries. The strategy underlines the shared interest of OECD and partner countries in maintaining the high quality of PISA in a way that is sustainable and affordable. It also emphasises the importance of engaging with OECD's five Key Partners and sets out proposals for strengthening the engagement with these countries.

48. The participation of Key Partners in PISA up to now can be summarised as follows:

- Brazil has participated in each round of PISA since its beginning in 2000. Brazil has Participant status in the PGB.

- China: Five provinces and municipalities have agreed to participate as adjudicated regions in PISA 2015 and three other regions, all of which have participated in previous cycles, will continue as individual participants in PISA 2015. China formally joined PISA in 2013.

- India: Two states participated in PISA 2009+, an extension of the PISA 2009 assessment.

- Indonesia has participated in each round of PISA since its beginning in 2000. Indonesia has Participant status in the PGB.

- South Africa: To date there has been no participation from South Africa in PISA.

49. PISA already covers several developing countries, but certain barriers remain in seeking to further broaden the global coverage of the assessment: the contextual information collected through the questionnaires needs to fit the actual student and school contexts in a wider range of countries, the assessment needs to provide a sufficiently fine resolution of skills at the lower end of the proficiency distribution, and some countries may need to build up the institutional capacity to implement large-scale international assessments.

50. The PISA for Development project has been introduced to seek to overcome such barriers for the participation of developing countries in PISA. The project aims to develop PISA survey instruments that are better adapted to the contexts found in developing countries, but which produce results at the same scale as the main PISA assessment, and to pilot these instruments among a number of countries. Easier versions of the mathematics and science tests may be developed drawing initially on the existing item pool, and the assessment of component reading skills that has already been implemented in PISA can be used to provide more detailed information of the reading skills at the lower end of the skills range. The project also aims at adapting the student and school questionnaires to fit the contexts of developing countries, for example in terms of students' socio-economic backgrounds and the resource challenges facing schools. Once developed and piloted, these additional instruments could feed into the development of the main PISA instruments.

51. Other on-going developments of PISA may also help to overcome barriers for broader participation in PISA. Although developing countries' capacity to implement computer-based assessments is still limited, assessments delivered on computer or other electronic devices may in the future allow for

more flexibility between countries and thus potentially open up for increased participation. Computer-based adaptive testing, as mentioned above, is a way to potentially expand the coverage range of the cognitive test without reducing the measurement efficiency at the ability levels currently covered by the test.

Objective 5

To maintain PISA as a system-level assessment, while recognising countries' desire to provide feedback to schools and facilitating countries to provide such feedback

52. PISA's main output is comparable system-level data on the quality, equity and efficiency of school systems. This principle is reflected in all elements of the survey, from the assessment frameworks to the design of the instruments to the sampling and survey procedures and to the analysis and reporting of the survey results. The systems covered by PISA are mainly national school systems, but also increasingly regional and other sub-national systems.

53. Some countries also provide school-level feedback to participating schools, in many cases as a means to encourage and acknowledge the schools' participation in the survey. Providing feedback to schools and the extent of such feedback is a national decision and the desire to do so varies from country to country. Over the last years, countries have been sharing their experiences of providing feedback to schools and an expert paper, discussing different ways to provide such feedback and the methodological challenges involved with it, has been prepared by an international expert.

54. The challenges to providing useful feedback on school-level results include the accuracy that can be provided given the relatively small student samples at each school, the delay from data collection to the publication of results, and the way the age-based samples in PISA split across grades in many countries. Countries may address the issue of timing by providing schools with preliminary results based on average item scores at the school level in the context of similar country-level scores. The calculation of average scores could be facilitated by providing countries with generalizable software or syntax files that would operate on the raw data files to create the item and school aggregates. Templates for reporting the results could also be prepared internationally and offered to countries to facilitate the production of reports. The school mean comparisons could then be redone using the plausible values from the data set when the international scales are ready. However, even at this point the relatively small student samples at each school will in many cases generate large error margins, in particular if the results need to be interpreted at the grade level to be useful. Countries that wish to provide useful feedback to participating schools may therefore need to consider increasing sample sizes at the school level beyond the PISA minimum standards.

55. Another way to facilitate countries to provide feedback to schools is by continuing to share experiences across countries and to further study the methodological challenges and the ways to resolve the challenges. It would be relevant for such further studies to include the possible implications to PISA in general of providing participating schools with results.

56. For countries that decide to implement the PISA-Based Test for Schools, this option will be another way of providing information to schools about their performance in an international context. The assessment is specifically designed to yield reliable results at the school level rather than provide aggregate results at the system level. However, PISA and the PISA-Based Test for Schools are two separate assessments and the latter does not in itself address the demand for providing feedback to schools as a means to encourage and acknowledge their participation in the main PISA assessment.

Objective 6

To seek ways to widen access to PISA for students with disabilities and other special education needs

57. PISA aims to be as inclusive as possible and all participating countries make efforts to maximise the coverage of 15-year-old students in their samples. The sampling standards permit up to a total of 5% of the relevant population to be excluded. In PISA 2009, all but five countries met this standard, and in 36 countries and economies the overall exclusion rate was less than 2%.

58. However, because of technical and methodological challenges, PISA has to date offered only limited accommodations for students with disabilities or other forms of special needs. As a result, some students must be excluded from the assessment, and in some countries exclusion rates are growing as more and more students are identified as having disabilities that require accommodations when the students are tested.

59. A concern with providing accommodations is that students' needs may vary to an extent that makes it difficult to address individual needs while guaranteeing comparability within and between countries. The accommodations selected should maximize on the following criteria: address as many needs as possible, be frequently used in participating countries, pose little or no threat to validity, be practical, inexpensive, and centrally developed. Exploratory studies undertaken by the Strategic Development Group suggest the use of 9 accommodations: graphic modifications, audio presentation, bilingual dictionary, dictation of answers, word processor, individual or small group settings, adaptive furniture or tools, extended time and rest periods.

60. A second concern is that some accommodations may appear to place students at a comparative advantage to those not using the accommodations. A solution to this may be to exclude students receiving accommodations from countries' samples. However, if large proportions of students accepted the accommodations, then countries would see rising exclusion rates, rather than declining exclusion rates and such exclusion may be difficult to justify.

61. Technological and methodological advances may enhance the possibility for future PISA studies to provide for accommodations that could address some of the most important needs without placing accommodated students at a comparative advantage. These advantages have the potential to enable PISA to expand the accommodations available to students and thus enable more students to participate in PISA. PISA results would, in turn, better represent the target population, but just as important, greater availability of accommodations would enable PISA to address what is increasingly a legal requirement to include students who are capable of taking the assessment and can demonstrate what they know and can do.

62. Accommodations that have been allowed in PISA to date include, for example, small-group or one-on-one testing. However, the testing accommodation most frequently required by students with special education needs is unlimited or extended time, and that is not permitted for students taking the standard PISA test booklets. There are also other accommodations for students with some forms of special needs that have not been widely allowed in PISA, such as, for example, large print, use of special fonts or magnification and assistance inputting responses (*e.g.*, allow students to orally deliver their responses into a microphone). Without these types of accommodations students who could otherwise participate in a meaningful way must be excluded from PISA.

63. While the move towards computer delivery may mean that it will become easier to accommodate students with some forms of special needs, it may also pose a special challenge for students with some learning difficulties (for example, complex webs of hyperlinked pages with multiple graphs and figures can be especially challenging for students with dyslexia while software that reads out loud the material and

allow texts to be displayed using special fonts may be of help for the same group of students). PISA could investigate the feasibility and advisability of several potential accommodations. For example, students who require unlimited or extended time could be allowed to take that extra time, but PISA could "use" only the data that is captured within the two hours allowed for the assessment. Technological and methodological innovations create new opportunities with respect to testing students with special needs, which, essentially would revolve around the issue of ensuring a more tailor-made approach to assessment instrument delivery.

64. A second area where computer delivery could possibly lead to enhancing PISA's capacity to aid education policies regarding inclusive education is by allowing more flexibility in possibilities for languages in PISA, for example in bilingual situations.

65. PISA may benefit from exploring how inclusiveness can be increased through additional accommodations while ensuring that test administrations remain sufficiently standardized to enable comparisons across students and countries. This could be accomplished by investigating how technology could be used to increase inclusiveness and then conducting small-scale trials, perhaps embedded in a field test administration, to determine how various accommodations could be made available and what impact their availability would have on exclusion rates. As far as countries differ in their provision of accommodations, it may also be considered to report what provision is made in each participating country and provide guidance considering the implications for how the data should be interpreted in that light.

REFERENCES

Binkley, M., Sternberg, R., Jones, S. and Nohara, D. (1999) An overarching framework for understanding and assessing life skills. In *Frameworks: Working Drafts.* Briefing materials for Adult Literacy and Life skills Survey. National Study Managers Meeting, Luxembourg.

Rychen, D.S. (2003) Key competencies: Meeting Important Challenges in Life. In Rychen, D.S. and Salganik, L.H. *Key Competencies for a successful Life and Well-functioning Society*. Hogrefe and Huber, Göttingen.

Trier, U.P. (2003) Twelve countries contributing to DeSeCo: A summary report. In D.S. Rychen, L.H. Salganik and M.E. McLaughlin (Eds.) *Selected contributions to the 2nd DeSeCo Symposium.* Neuchâtel, Swiss Federal Statistics Office.

Wu, M. (2010), "Comparing the Similarities and Differences of PISA 2003 and TIMSS", *OECD Education Working Papers*, No. 32, OECD Publishing. http://dx.doi.org/10.1787/5km4psnm13nx-en