PISA 2018 INTEGRATED DESIGN

Doc.:
CY7_NPM(1603)02b_GEN_IntegratedDesign_1.docx

[NOTE: A draft version of this document was shared with the PGB in Munich. This updated version includes the assessment of Financial Literacy and final decisions regarding international options.]

First Meeting of the PISA 2018 National Project Managers

14-18 March 2016

Prague, Czech Republic

Produced by ETS, Core A Contractor

Contents

PISA 2018 Cognitive Assessment Design 6
Main Survey Assessment Design 7
Main Survey CBA Assessment Design 7
Main Survey PBA Assessment Design 10
Field Trial Assessment Design 11
Field Trial CBA Assessment Design 13
Field Trial PBA Assessment Design 16
Financial Literacy Design 17
Main Survey Design for Financial Literacy 17
Field Trial Design for Financial Literacy 17
PISA 2018 Questionnaire Design 19
Main Survey Questionnaire Design 19
Field Trial Questionnaire Design 20
Annexes 22
Annex A. Field Trial Computer-Based Integrated Design 22
Annex B. Field Trial Paper-Based Integrated Design 23
Annex C. Field Trial Paper-Based Integrated Design – UH Design 23

PISA 2018 INTEGRATED DESIGN

1. The purpose of this document is to outline the assessment and questionnaire design for PISA 2018. The PISA 2018 design will be built upon the major design and methodology innovations introduced for PISA 2015, which included a reconceptualization of minor domains to diminish differences in domain coverage across cycles, a linking study to evaluate and control for potential mode effects, and computer administration as the primary mode of delivery for all core domains. In addition, while multistage adaptive testing for newly developed tasks was in discussion for the 2015 design, a more fully adaptive assessment is part of the design for 2018.

2. Among other things, these design changes increased the number of items for each of the minor domains, allowing for more complete construct coverage and for a new methodological approach to be employed. In contrast to previous cycles where scaling was conducted separately for each cycle, with a subsequent equating to previous results through a single transformation, the methodology implemented from 2015 on will incorporate data from the relevant prior rounds in the nine-year domain rotation. This innovation informs trend item performance for scaling and analysis using a much larger database, thus providing a solid base for linking across cycles and between paper- and computer-based administrations on all scales. Taken together, these design and methodological innovations improve comparability across countries, stabilize parameter estimation and the measurement of trend, and improve the reliability of the inferences made from the data.

3. The current design paper is based on previous outlines prepared for the PISA Governing Board (PGB) for the 2015 assessment and discussed by the PISA Technical Advisory Group (TAG) in June 2015 for the 2018 cycle. The current paper was augmented by additional information that is specific to the 2018 cycle. As part of the 2018 design, Global Competence, the new innovative domain for 2018, will be fully integrated into the assessment design together with the core domains of Reading, Mathematical Literacy, and Scientific Literacy. Furthermore, as the contractor for Core A, we will work with the other contractors for 2018 to develop and implement a fully integrated design that takes into account the priorities for context questionnaires and sampling.

4. The Field Trial design in preparation for the 2018 Main Survey includes complexities that are due to PISA being a computer-based survey that continues to emphasize innovation and advances in the methodologies and procedures used for each cycle. The 2018 cycle must continue to collect reliable, valid, and comparable information about a widening range of knowledge, skills, and context factors, which will be facilitated by the proposal made to the PGB and Strategic Development Group (SDG) to introduce multistage adaptive testing in the 2018 Main Survey. In addition, for a subset of countries, PISA 2018 will continue to offer a non-adaptive, trend-only paper-based version, which will link to past cycles while not containing any of the innovative domains and interactive features in tasks available on the computer.

5. This document presents the proposed designs for the cognitive assessment as well as the questionnaires for both the Main Survey and the Field Trial. However, detailed information will be presented for the Field Trial only, as the final design of the Main Survey will depend in part on the evaluation of the Field Trial data. The designs presented here are based on the following assumptions:

- The design includes the assessment of Reading, Mathematical, and Scientific Literacy as well as Global Competence for every participating country. Those choosing the paper-and-pencil option will

be limited to assessing Reading, Mathematical, and Scientific Literacy trend items. No new cognitive items will be developed for paper and pencil. Regardless of which mode of delivery a country chooses, this portion of the survey will take approximately 120 minutes to complete.

- As the Field Trial needs to be conducted in the same mode as the Main Survey, all countries planning to implement PISA as a computer-based assessment (CBA) in the Main Survey will also do so in the Field Trial.

- Paper-based instruments linking to previous surveys will be provided for countries unable or unwilling to test their students by computer and will be limited to the set of trend items established in 2015 for the three core domains.

- Computer-based trend clusters for Mathematics and Reading will be identical to those used in 2015. Science trend clusters will comprise units included in the trend clusters along with newly developed units in 2015, resulting in broadened trend materials for computer-based countries for the 2018 Field Trial and Main Survey. Paper-based Assessment (PBA) countries in 2018 will continue to rely on the 2015 PBA trend clusters.

- The workflow for the cognitive assessment and student context questionnaires will consist of separate components to accommodate various options: i) a cognitive testing session of 120 minutes, ii) a student questionnaire session of approximately 35 minutes; iii) optional questionnaires for students that include an Educational Career Questionnaire (10 minutes), an Information and Communications Technology [ICT] Familiarity Questionnaire (10 minutes) and a Well-Being Questionnaire (15 minutes), and iv) an optional computer-based assessment of Financial Literacy (120 minutes). In addition, v) a School Questionnaire (45 minutes) will be administered to school principals; vi) an optional Parent Questionnaire (PQ) (30 minutes) and/or vii) an optional Teacher Questionnaire (TQ) (30 minutes) could be administered for countries selecting these options.

- An optional assessment of Financial Literacy will be incorporated into the survey design. Per request of the OECD, Financial Literacy will be administered through an additional sample of students beyond the 6,300 students. These students will take one hour of selected math and reading clusters and one hour of Financial Literacy. Data for the Financial Literacy sample will be analyzed separately from the regular PISA data, and a set of plausible values will be produced that can be used to study relationships among these three domains.

6. Figure 1 shows simplified versions of the Integrated Design for PISA 2018 for both students as well as non-student components.

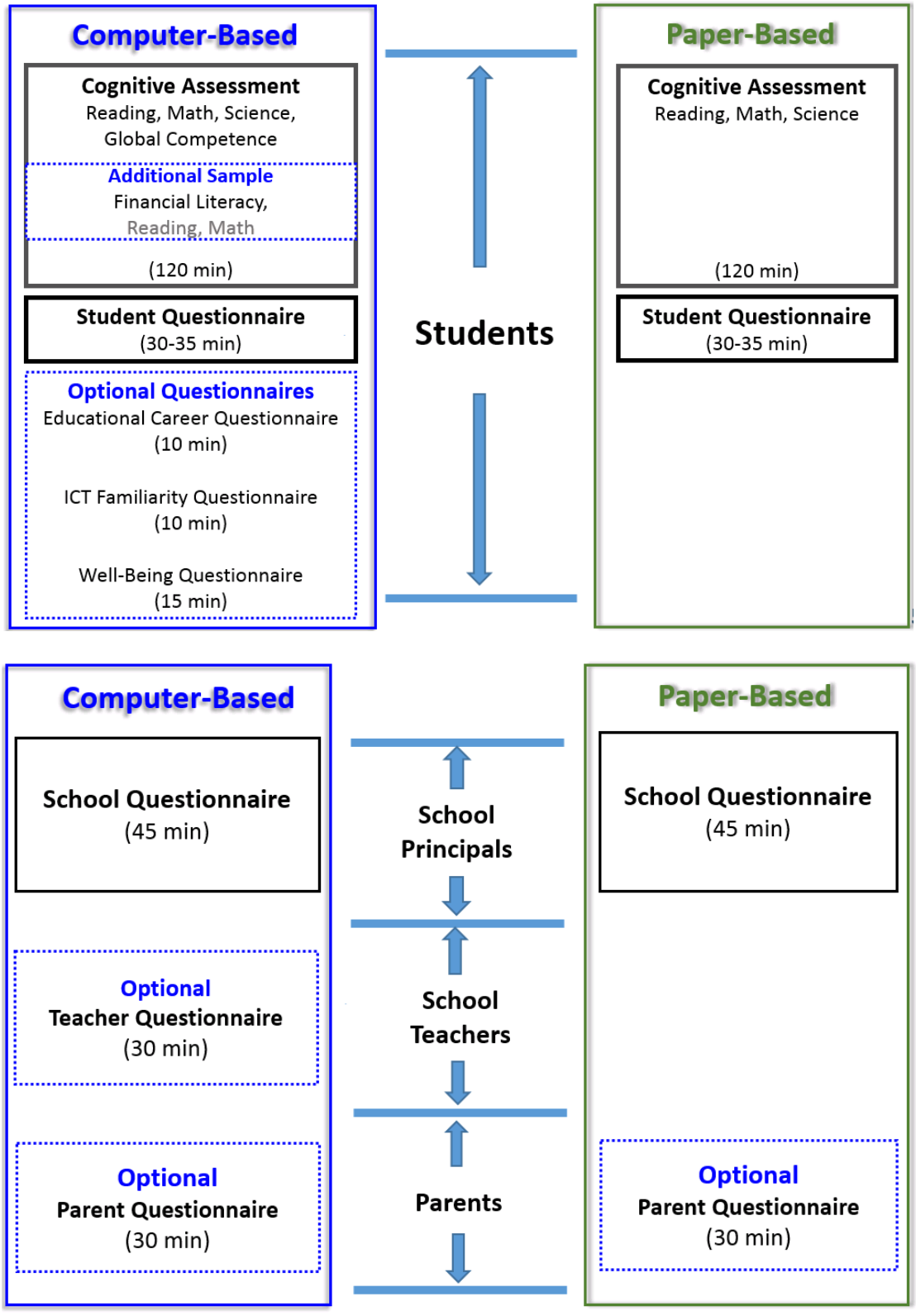


Figure 1. PISA 2018 Integrated Design (Simplified)

PISA 2018 Cognitive Assessment Design

7. Under the cognitive assessment design for PISA 2018, the total testing time for measuring the four domains of Reading, Mathematical, and Scientific Literacy, as well as Global Competence will remain at two hours (120 minutes) for each student. While computer-based delivery will remain the primary mode for 2018, paper-and-pencil instruments linking to earlier PISA cycles will be provided for countries that are not able or willing to test their students by computer. However, a return to paper and pencil will not be considered for countries that already transitioned to CBA in 2015. The paper-based assessment will be limited to existing Reading, Mathematical, and Scientific Literacy items administered as trend in 2015. No changes in PBA trend clusters will be implemented.

8. The designs that will be described in detail in the following sections cover the four skill domains that will be assessed in 2018. The designs follow the 2015 PISA design that increased the construct coverage and reduced design distinctions between minor and major domains, hence improving the information that PISA can provide to policy makers concerning the distribution of skills in student populations. In addition, these assessment designs will use advanced measurement and computer technologies to deliver PISA in an efficient manner.[1] In summary, the PISA 2018 assessment design will provide participating countries with the following information:

- population distributions in Reading Literacy, which reflect the new 2018 framework for this major domain, as well as links to the framework and scale developed in 2009 with CBA trend;
- population distributions in Mathematical Literacy linked to the 2012 framework and the Mathematical Literacy scale;
- population distributions in Scientific Literacy linked to the 2015 framework and the Scientific Literacy scale;
- population distributions in Global Competence;
- pairwise covariance estimates among each of the four domains; and
- three-way association estimates among the four cognitive domains, including the three core PISA domains (Reading, Mathematical, and Scientific Literacy).

9. Table 1 shows the number of clusters for each domain and mode that will be included in PISA 2018 Field Trial and Main Survey.

[1] We expect to introduce multistage adaptive testing for the Main Survey in PISA 2018. The extent to which this type of adaptive testing will be introduced will depend on decisions that need to be made regarding the use of intact clusters. These issues have been discussed with the PGB and SDG, and their recommendation to move forward with our proposal was discussed with the TAG that supported the Field Trial design to investigate the effects of variable unit position within the cluster in preparation for multi-stage adaptive testing in 2018.

Table 1. Domain coverage for PISA 2018

Domain Coverage	NEW items (CBA only)		TREND (CBA and PBA)	
	Field Trial	**Main Survey**	**Field Trial**	**Main Survey**
Reading Literacy	12 30-min clusters	9 30-min clusters	6 30-min clusters	6 30-min clusters
Mathematical Literacy	NO	NO	6 30-min clusters	6 30-min clusters
Scientific Literacy	NO	NO	6 30-min clusters	6 30-min clusters
Global Competence	4 30-min clusters	3 30-min clusters	NO	NO

10. It is important for countries to note that the Contractors will assume full responsibility for assembling all paper and computer forms of the cognitive instruments. While there are many more forms in PISA 2018 than in previous cycles up to 2012, there will be no additional work on the part of the country in assembling these forms. Trend clusters for the Mathematics and Reading domains are identical to those used in PISA 2015. Six new computer-based trend clusters for Scientific Literacy will be assembled from the item pools of six computer-based trend and six new computer-based clusters used in 2015. It should be noted that the paper-based trend clusters for Scientific Literacy will remain the same as those assembled for 2015.

Main Survey Assessment Design

11. The Main Survey assessment design for PISA 2018 covers the domains of Reading, Mathematical, and Scientific Literacy for both computer- and paper-based modes. The computer-based design adds the fourth domain of Global Competence as the innovative domain for PISA 2018. These designs require participating countries to draw a random sample of at least 150 schools representing their national population of 15-year-old students. Countries taking the CBA will need to sample 42 students from each of 150 schools for a total sample of 6,300 students, while countries taking the PBA will need to sample 35 students from each of 150 schools for a total sample of 5,250. It is important to understand that 88-92% of students will receive a test form that consists of four 30-minute clusters (or sets of tasks) assembled from two domains, resulting in one hour of assessment time per domain. An additional 8-12% of students will receive forms that consist of four 30-minute clusters (sets of units) covering three of the four core domains. Two clusters measuring Reading Literacy are included in each of these forms.

Main Survey CBA Assessment Design

12. The 2015 PISA design supported the expressed long-term goal of diminishing the differences between minor and major domain coverage. In addition, the introduction of the computer-based mode of assessment ensures that the design innovations can be implemented with the required spiraling of the augmented set of trend clusters. These innovations introduced in 2015 form the basis for the 2018 assessment, which continues to capitalize on the greater stability of estimates achieved by the increased construct coverage for minor domains.

13. For the 2015 design, adaptive testing was considered but suspended in light of the need to manage both the design innovations introduced at that time as well as the changes in delivery mode. Also, one of the design criteria required in the call for tender for 2015 was the use of intact 30-minute clusters, which prevented the use of an adaptive design. In order to move beyond this restriction, it is imperative to verify the invariance of item parameters under conditions that break up 30-minute clusters of units into smaller sets of items presented in various positions and combinations. This invariance is a necessary precondition for the introduction of an adaptive design. For the new 2018 design, the Field Trial will provide information that can be used to help evaluate this prerequisite. This issue will be further addressed in the section on the Field Trial design.

14. Depending on the results of the invariance analysis to be conducted on the Field Trial data, a Main Survey design that includes adaptive testing will be developed that minimizes any effects of adaptivity. In the unlikely event that the Field Trial indicates intolerable violations of parameter invariance due to unit position effects, we will develop and implement a non-adaptive design for 2018.

15. The proposed 2018 design shown here comprises units that include the six trend clusters from Reading and Mathematical Literacy implemented in the 2015 cycle. Six trend clusters for Scientific Literacy will be assembled from the six trend clusters and six new clusters used in 2015, as shown in Table 1. The six clusters of Scientific Literacy tasks will carry not only the trend information but also link to the new items developed to reflect the 2015 framework. In addition, three clusters of Global Competence (GC) items will be assembled for the Main Survey. These clusters across the four domains will be organized into test forms as shown in Figure 2.

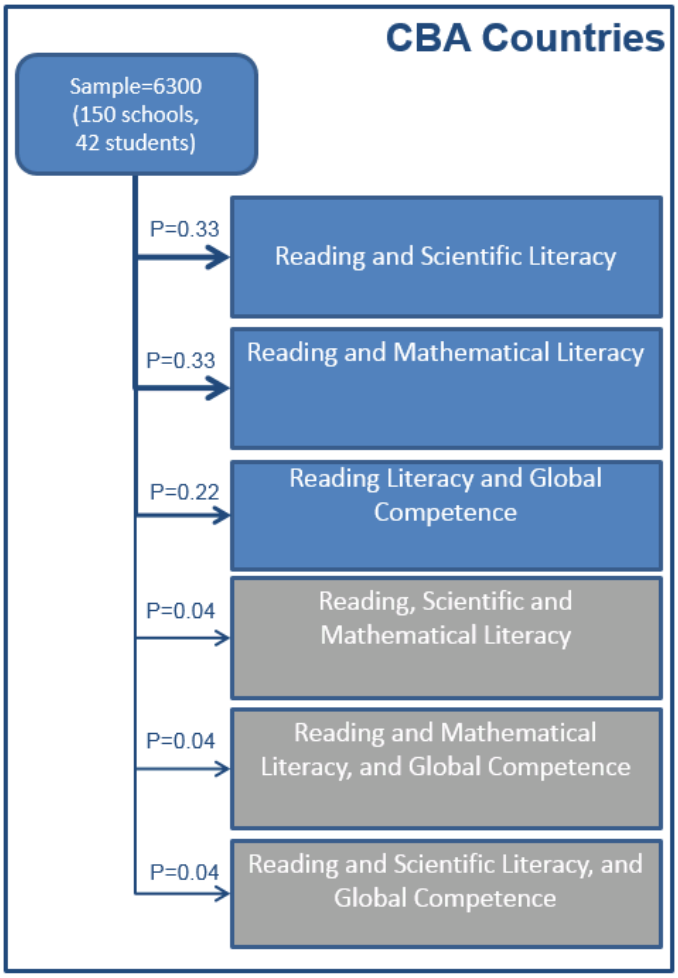


Figure 2. Main Survey Computer-Based Assessment Design

16. As reflected in Figure 2, there are multiple forms in each of three different kinds of test forms that represent various combinations of two of the four domains – 88% of students receive one of these forms according to a prescribed probability. These combinations include: i) Reading and Scientific Literacy, ii) Reading and Mathematical Literacy, and iii) Reading Literacy and Global Competence. In these test forms, students take one hour of Reading Literacy plus one hour of a minor domain – Scientific Literacy, Mathematical Literacy, or Global Competence. These forms will be sampled at a higher rate and provide the necessary covariance information between Reading Literacy and each of the three minor domains.

17. In addition, there are other forms that will provide tri-variate information about three of the four domains, sampled at a lower rate so that only 12% of students receive one of these forms. In these forms, students receive one hour of Reading Literacy plus two 30-minute clusters of items from each of the other three domains. These combinations are: iv) Scientific, Reading, and Mathematical Literacy; v) Reading Literacy, Mathematical Literacy, and Global Competence; and vi) Reading Literacy, Scientific Literacy, and Global Competence. It is important to note that these three-domain test forms will ensure that covariance estimates among the four domains are based on the joint assessment of two as well as three domains.

18. As Reading Literacy is the major domain for 2018, it is present in all test forms and paired with one or two of the other three domains, and each of the different combinations of domains is balanced in terms of position. The design also reflects the fact that the random assignment of a form within a school follows a specific preassigned probability distribution. According to this design, 33% of students within each school will be assigned to one of the Reading and Scientific Literacy test forms. Another 33% will be assigned to one of the Reading and Mathematical Literacy test forms. In addition, some 22% of the sampled students within each school will be assigned to one of the Reading Literacy and Global Competence test forms. To provide additional covariance information, 4% of students will be assigned to one of the Reading Literacy, Mathematical Literacy, and Global Competence test forms; 4% to one of the Reading Literacy, Scientific Literacy, and Global Competence test forms; and 4% to one of the Reading, Mathematical, and Scientific Literacy test forms.[2] The specific form of the adaptive testing design proposed for the 2018 assessment is commonly referred to as multistage adaptive testing. This approach is particularly well suited for tests that consist of units that in turn are composed of multiple items, some of which may require human coding. In addition, multistage adaptive testing utilizes routing decisions that are based on performance on a set of items, and hence can be expected to be more robust against item-by-country interactions than item-level adaptive algorithms.

19. The multistage design that is being recommended for PISA 2018 was successfully used in the OECD Programme for International Assessment of Adult Competencies (PIAAC). The design shown in Figure 2 can accommodate a multistage adaptive testing design easily for the main domain as all students will be assessed for an hour. For the minor domain, a more limited version of adaptive testing involving fewer stages will be implemented. Additional information relating to adaptive testing for PISA 2018 can be found in the document *PISA 2018 Computer Adaptive Testing: CY6_PGB(1503)_GEN_PISA2018_Computer Adaptive Testing*, which was discussed at the March 2015 PGB meeting.

Main Survey PBA Assessment Design

20. As was the case in PISA 2015, countries that are not able or willing to use computer-based delivery for PISA 2018 can choose to test their students using a paper design. This assessment will be very similar to the PBA design used in 2015. It will consist of paper-based trend items for Scientific, Reading, and Mathematical Literacy that rely heavily on the frameworks and assessments that were developed when each was a major domain – 2006 for Scientific Literacy, 2009 for Reading Literacy, and 2012 for Mathematical Literacy. This PBA design is shown in Figure 3.

[2] These percentages are based on random assignment of test forms to students across schools. Each student in each classroom has a real probability of receiving any of the forms.

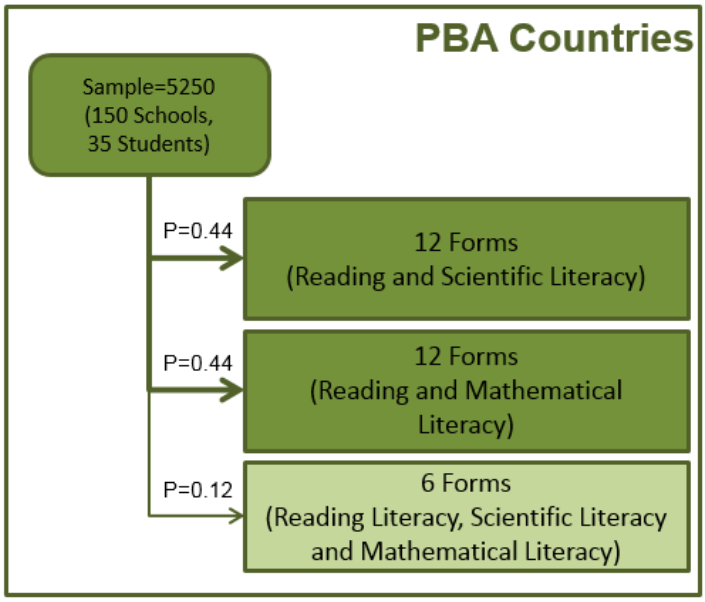


Figure 3. Main Survey Paper-Based Assessment

21. As with the CBA design shown in Figure 2, the PBA design includes six intact clusters from each of the three core domains (see Table 1 as well). These intact clusters will provide trend information for Mathematical, Reading, and Scientific Literacy. It is important to note that in the paper-based survey, the assessment of Reading Literacy will be limited to trend items and identical to the 2015 PBA items. It will not include any new items based on the new 2018 Reading Literacy framework as those items are developed for the computer-based platform only. As with the previous designs, there are several different sets of forms that combine two or three of the domains to provide covariance information. These combinations include: i) Reading and Scientific Literacy; ii) Reading and Mathematical Literacy; and iii) Scientific, Reading, and Mathematical Literacy. The position of items within each domain will be balanced and the assignment of a form within a school follows a probability based sampling of forms like the one used for assigning test forms in the CBA design. In this PBA design, 44% of students will be assigned to one of the 12 Reading and Scientific Literacy forms and another 44% will be assigned to one of the 12 Reading and Mathematical Literacy forms. The remaining 12% of students will be assigned to one of six Scientific, Reading, and Mathematical Literacy forms.

22. We will continue to offer the UH options that were available in 2015. The mode will be the same as for the main booklets, that is, CBA countries will have a CBA-based UH booklet, while PBA countries have a PBA UH booklet if utilized.

Field Trial Assessment Design

23. This section illustrates the Field Trial design for PISA 2018, which is necessarily complex because it must provide evidence to support various goals and inferences in preparation for the Main Survey. One of the goals to be met by the Field Trial is collecting information in preparation of the planned introduction of multistage adaptive testing. In order to examine the feasibility of implementing multistage adaptive testing for the Main Survey, the Field Trial will be used to gather information about testing units in variable instead of fixed positions within 30-minute (intact) clusters. In other words, we will test the notion that item parameter invariance is only supported when using intact clusters. The observation of block order

effects that triggered this notion in early PISA rounds has led to the assumption that intact cluster positions are needed for parameter invariance to hold. This is an observation that was unique to PISA and was not found in other large-scale assessments, including PIAAC and the National Assessment of Educational Progress (NAEP). There was no need to adjust item parameters based on relative item positions in the cognitive instruments. In addition, a rescaling study conducted on the joint database of all historical PISA data collected between 2000 and 2012 showed good stability of item parameters overall across multiple survey cycles even though over time there were deviations from the strict application of the "intact cluster" paradigm. The 2018 Field Trial will provide additional information in regard to item parameter invariance under variable unit positions.

24. For the 2018 PISA Field Trial, we will consider a unit to represent the minimum granular size of item sets at which adaptiveness can take place. Units are sets of items under a common stem that can be considered as the organizing grain size aligned randomly or guided by adaptiveness. While within-unit adaptiveness would be possible in principle, altering the sequence of items within a unit could change the substantive meaning or context, so no variations will be introduced within a unit. However, the sequence of units within a cluster can be changed to examine parameter invariance relative to unit position. Examining and ensuring parameter invariance at the unit level is a necessary condition for PISA 2018 to move to adaptive testing (*PISA 2018 Computer Adaptive Testing: CY6_PGB(1503)_GEN_PISA2018_Computer Adaptive Testing*).

The goals of the Field Trial design include:

- evaluation of the invariance of item parameters compared to previous PISA cycles for the 2018 cycle (CBA and PBA);
- evaluation of the invariance of item characteristics in regard to the size of intact item sets; that is, a comparison of 30-minute clusters found in prior PISA rounds versus varying positions of smaller collections of units in preparation for adaptive testing;
- obtaining initial item parameters for the evaluation of new Reading Literacy and Global Competence items, and for the selection of a final set of items used in the Main Survey for these new units;
- evaluating sampling and survey operations; and,
- assessing how well the computer platform functions within and across participating countries.

25. Because the primary goal of the Field Trial is to support the goals noted above and not to estimate the proficiency distribution of national populations, the sampling requirements differ from those specified in this paper for the Main Survey.

26. Like the Main Survey design, the Field Trial design for PISA 2018 implements one CBA design including Reading, Mathematical, and Scientific Literacy and Global Competence and one PBA design that involves the core domains of Reading, Mathematical, and Scientific Literacy as they were implemented in 2015.

27. The recommended design for countries choosing computer delivery is to sample a minimum of 28 schools for the Field Trial. Within each of these schools, 71 students will need to be selected and tested for a target sample size of 1,992 students. The preferred number of students to be sampled from

each school participating in the Field Trial for 2018 is similar to 2015. While this is an increase from previous cycles by 20 students, the Field Trial in 2018 will again serve the purpose of comparing several design variants, this time with respect to adaptive testing. Selecting schools having a large eligible student body and with sufficient computer facilities should be considered. Although, the PISA 2018 Field Trial design is more adaptable than the previous design in terms of number of schools, increasing the number of schools while sampling a reduced per-school student group is not recommended because it would make comparisons among the results from each group more difficult. For countries having difficulty in finding large schools, the following design can be considered: 50 students from each of 39 sampled schools randomly divided into three groups of 15, 17, and 18 students, respectively. Random assignment of forms within a group is required. This design would yield 1,950 cases. We recognize that some countries may choose to modify these design requirements to meet unique conditions present in participating countries. If a country chooses to use a different design, it is important to choose one that is best suited to its needs while supporting the goals of the Field Trial.

28. All students will be randomly assigned to respond to one of three groups: CBA Trend, CBA Trend/New Reading, and CBA New Reading and Global Competence. It is important to understand that each student will receive one of the forms from his or her assigned group and that each test form will consist of four 30-minute clusters assembled from at most two domains, resulting in at least one hour of assessment time per domain, with a total of two hours of testing time per student.

29. The forms associated with each group have specific functions that support one or more of the Field Trial goals listed earlier. It is important to note that the analyses to be carried out on the Field Trial data are typically geared toward aggregated data across countries. The Field Trial data will provide evidence on the level of reporting scale performance and item invariance rather than providing information at the country level. The Field Trial sample size is not large enough to evaluate country-level results separately.

30. Countries that choose to measure student performance with only paper-and-pencil forms will have a much-reduced sample size because they will be using forms from previous cycles. The goals for these countries will be mainly focused on testing operations and data-related procedures. These countries will also need to sample 25 schools but only have to select 36 students from each school for a total Field Trial sample of 900.

Field Trial CBA Assessment Design

31. Three randomly equivalent groups will be formed within a school or across multiple schools. These groups will be administered specific forms to meet evaluative goals noted above, are shown in Figure 4 and described below.

Group 1 – CBA Trend, Fixed Unit Order (FUO) (Intact Cluster)

32. The Field Trial CBA design includes six intact trend clusters from Reading and Mathematical, Literacy identical to those assembled in 2015 and used as such previously. In addition, the six Scientific Literacy clusters will be assembled from trend and new computer-based Science units used in the 2015 Main Survey thus resulting in a mixture of trend and new Science units. So, strictly speaking, there are also no “intact” clusters for Science compared to past assessments. The data from this group provides information with regard to the degree of invariance of psychometric characteristics from 2015 to 2018 in Reading and Mathematics. For Science, it is possible to examine the variability of psychometric characteristics from 2015 to 2018 and the impact of different unit order in 2018. The total sample size for

this group, referred to as Group 1-CBA Trend, includes Forms 1-18 and is designed to be 576 to yield an expected number of 128 responses per item per country.

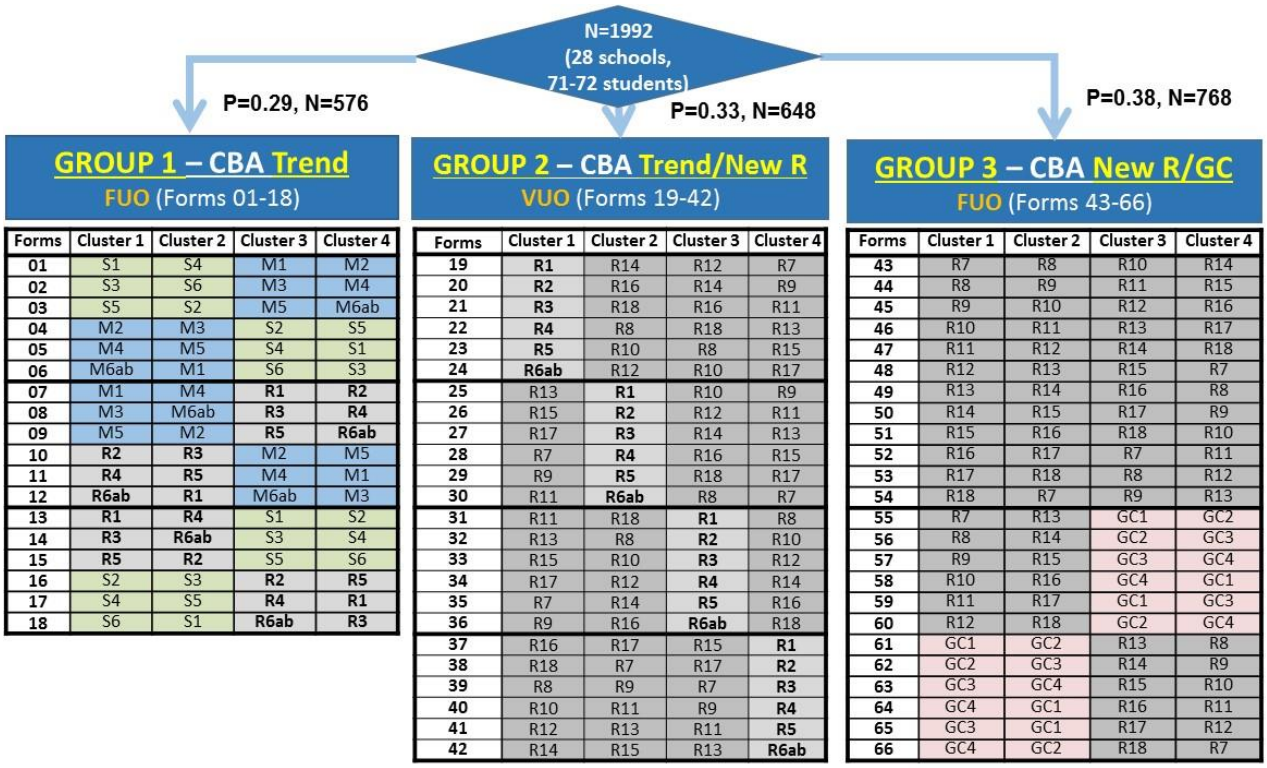
Group 2 – CBA Trend and New Reading, Variable Unit Order (VUO)

33. This group, referred to as Group 2 – CBA Trend and New Reading, includes 24 forms (Forms 19 – 42) and is designed to provide variations of unit orders within clusters. Unit positions within a cluster will vary based on a limited number of sequences to examine the invariance of psychometric characteristics relative to the Group 1 results (Fixed Unit Order). Each of the 24 forms consists of a combination of one of six trend Reading clusters and three of 12 new Reading clusters as shown in Figure 4. Every trend cluster will be paired with each new cluster once and appears in all four cluster positions once or twice in each position (six times altogether). The total sample size for this group is designed to be 648 to yield, per country, an expected number of 108 responses per trend item and 162 responses per new item.

Group 3 – CBA New Reading and Global Competence, Fixed Unit Order

34. Clusters in this group, referred to as Group 3 – CBA New Reading, are based on a fixed order of units to provide a basis for comparisons with variable unit order forms administered in Group 2. There are two sets of 12 forms each in this group. The first set of 12 forms (Forms 43-54) is assembled from new Reading clusters and each cluster is presented once in every position and paired with other clusters at least once. The second set of 12 forms (Forms 55-66) includes two new Reading clusters and two Global Competence clusters. Cluster positions are balanced for both domains. Every form will be administered to 32 students, thus these two sets will be administered to 768 students altogether (384 students in each set).

35. As shown in Figure 4, the Field Trial CBA Design with Global Competence is based on sampling 72 students per school. Students will be randomly assigned to one of these three groups. According to this design, 21 of the 72 students (or 29%) will be assigned to one of the forms in Group 1-CBA Trend; 24 of the 72 students (or 33%) will be assigned to one of the test forms in Group 2-CBA Trend and New Reading; and 27 of the 72 students (or 38%) will be assigned to one of the test forms in Set 1 (Forms 43-54) of Group 3-CBA New Reading and Global Competence or Set 2 (Forms 55-66) of Group 3-CBA New Reading/Global Competence. It is important to remember that each test form will contain four 30-minute clusters for a total testing time of two hours per student.



Forms	Cluster 1	Cluster 2	Cluster 3	Cluster 4
01	S1	S4	M1	M2
02	S3	S6	M3	M4
03	S5	S2	M5	M6ab
04	M2	M3	S2	S5
05	M4	M5	S4	S1
06	M6ab	M1	S6	S3
07	M1	M4	**R1**	**R2**
08	M3	M6ab	**R3**	**R4**
09	M5	M2	**R5**	**R6ab**
10	**R2**	**R3**	M2	M5
11	**R4**	**R5**	M4	M1
12	**R6ab**	**R1**	M6ab	M3
13	**R1**	**R4**	S1	S2
14	**R3**	**R6ab**	S3	S4
15	**R5**	**R2**	S5	S6
16	S2	S3	**R2**	**R5**
17	S4	S5	**R4**	**R1**
18	S6	S1	**R6ab**	**R3**

Forms	Cluster 1	Cluster 2	Cluster 3	Cluster 4
19	**R1**	R14	R12	R7
20	**R2**	R16	R14	R9
21	**R3**	R18	R16	R11
22	**R4**	R8	R18	R13
23	**R5**	R10	R8	R15
24	**R6ab**	R12	R10	R17
25	R13	**R1**	R10	R9
26	R15	**R2**	R12	R11
27	R17	**R3**	R14	R13
28	R7	**R4**	R16	R15
29	R9	**R5**	R18	R17
30	R11	**R6ab**	R8	R7
31	R11	R18	**R1**	R8
32	R13	R8	**R2**	R10
33	R15	R10	**R3**	R12
34	R17	R12	**R4**	R14
35	R7	R14	**R5**	R16
36	R9	R16	**R6ab**	R18
37	R16	R17	R15	**R1**
38	R18	R7	R17	**R2**
39	R8	R9	R7	**R3**
40	R10	R11	R9	**R4**
41	R12	R13	R11	**R5**
42	R14	R15	R13	**R6ab**

Forms	Cluster 1	Cluster 2	Cluster 3	Cluster 4
43	R7	R8	R10	R14
44	R8	R9	R11	R15
45	R9	R10	R12	R16
46	R10	R11	R13	R17
47	R11	R12	R14	R18
48	R12	R13	R15	R7
49	R13	R14	R16	R8
50	R14	R15	R17	R9
51	R15	R16	R18	R10
52	R16	R17	R7	R11
53	R17	R18	R8	R12
54	R18	R7	R9	R13
55	R7	R13	GC1	GC2
56	R8	R14	GC2	GC3
57	R9	R15	GC3	GC4
58	R10	R16	GC4	GC1
59	R11	R17	GC1	GC3
60	R12	R18	GC2	GC4
61	GC1	GC2	R13	R8
62	GC2	GC3	R14	R9
63	GC3	GC4	R15	R10
64	GC4	GC1	R16	R11
65	GC3	GC1	R17	R12
66	GC4	GC2	R18	R7

Figure 4. Field Trial Computer-Based Assessment Design, with Global Competence

Where:

- R1-R6ab represent CBA trend Reading clusters
- R7-R18 represent CBA new Reading clusters
- M1-M6ab represent CBA trend Math clusters
- S1-S6 represent CBA trend Science clusters (assembled from 2015 trend and new items)
- GC1-GC4 represent CBA new Global Competence clusters

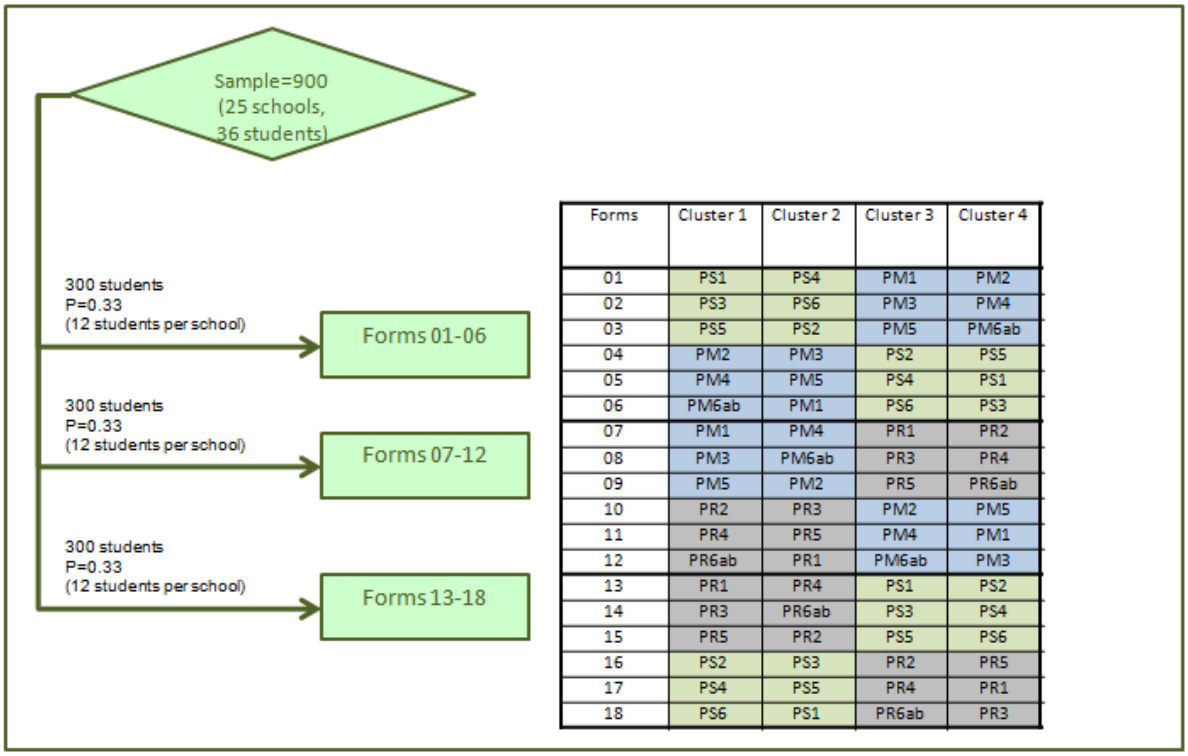
Note: FUO = fixed unit order; VUO = variable unit order

36. The introduction of the new design in PISA 2015 to reduce the distinction between major and minor domain is supported by analytic approaches that utilize the multi-cycle character of the PISA database. Data collected during the major domain cycle provides a basis for the analyses of two subsequent minor domain cycles. In terms of the Field Trial data analysis, Group 1 forms can be directly linked to the existing data from prior cycles, The variability in the psychometric characteristics of the 2018 Field Trial data collected relative to the 2015 analysis results gives a baseline for the magnitude of error we must expect across data collections when moving to adaptive testing. The variability in Group 1, given fixed unit order within a cluster with full construct coverage, is the lower bound that can be used to evaluate the variability of psychometric characteristics observed for Group 2 (variable unit order) and Group 3 (fixed unit order) of the new Reading clusters. Also the same variability will be referred to for comparisons of Reading trend clusters of Group 1 (fixed unit order) and Group 2 (variable unit order).

37. It should be emphasized again that we are examining Field Trial data at the scale level and making inferences of invariance of item characteristics with respect to unit order within a cluster on the aggregated data only across countries.

Field Trial PBA Assessment Design

38. Countries that choose the PBA design for the Main Survey will be measuring student performance with paper-and-pencil forms in the Field Trial. This paper-based Field Trial will have a reduced sample size compared to the CBA Field Trial because countries will be using only item clusters from previous cycles. While these countries will sample 25 schools, they will need to select only 36 students from each school for a total Field Trial sample size of 900. These students will be randomly assigned one of the 18 paper-and-pencil forms containing the trend items from two of the three core domains for PISA – Reading, Mathematical, and Scientific Literacy. This design is shown in Figure 5.



Forms	Cluster 1	Cluster 2	Cluster 3	Cluster 4
01	PS1	PS4	PM1	PM2
02	PS3	PS6	PM3	PM4
03	PS5	PS2	PM5	PM6ab
04	PM2	PM3	PS2	PS5
05	PM4	PM5	PS4	PS1
06	PM6ab	PM1	PS6	PS3
07	PM1	PM4	PR1	PR2
08	PM3	PM6ab	PR3	PR4
09	PM5	PM2	PR5	PR6ab
10	PR2	PR3	PM2	PM5
11	PR4	PR5	PM4	PM1
12	PR6ab	PR1	PM6ab	PM3
13	PR1	PR4	PS1	PS2
14	PR3	PR6ab	PS3	PS4
15	PR5	PR2	PS5	PS6
16	PS2	PS3	PR2	PR5
17	PS4	PS5	PR4	PR1
18	PS6	PS1	PR6ab	PR3

Figure 5. Field Trial Paper-Based Assessment Design

Where:

- R1-R6ab represent PBA trend Reading clusters
- M1-M6ab represent PBA trend Math clusters
- S1-S6 represent PBA trend Science clusters (same clusters from 2015)

Financial Literacy Design

39. The assessment of Financial Literacy will be offered again in PISA 2018 as an optional computer-based component. It will be based on a revised Financial Literacy framework based on the PISA 2012 Framework *[Doc. Ref.: FinLit_Frmwrk_PISA2012.doc]* that was used for PISA 2015. The cognitive instruments will include trend items from 2012/2015 plus a set of new interactive items that were developed specifically for PISA 2018.

40. Following a request from the OECD Secretariat, Financial Literacy will be administered to a separate sample of PISA-eligible students who take a combination of Reading, Mathematical and Financial Literacy tasks.

Main Survey Design for Financial Literacy

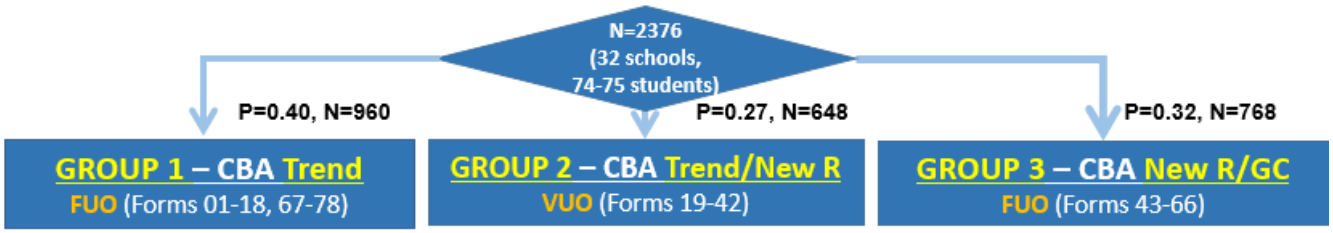
41. Under the computer-based assessment design for PISA 2018, the total testing time for measuring the three domains of Reading, Mathematical, and Scientific Literacy for the Financial Literacy sample will remain at two hours (120 minutes) for each student.

42. For the Main Survey, the Financial Literacy sample will be of 1,650 students, with each student taking 60 minutes of Financial Literacy tasks, 30 minutes of Mathematical Literacy tasks and 30 minutes of Reading Literacy tasks. Data for the Financial Literacy sample will be analyzed separately from the regular PISA data, and a set of plausible values will be produced that can be used to study relationships among these three domains.

Field Trial Design for Financial Literacy

43. For the Field Trial design, the Financial Literacy sample will be of 384 students who will each be assigned one of the 12 Financial Literacy testing forms. These forms include 60 minutes of Financial Literacy tasks, 30 minutes of Reading Literacy tasks and 30 minutes of Mathematical Literacy tasks. They are based on 3 clusters of Financial Literacy (FL1 to FL3) assembled from both 2015 trend items and 2018 new items in addition to the six trend clusters each Reading Literacy (R1 to R6ab) and Mathematical Literacy (M1 to M6ab). The 12 Financial Literacy forms – referred to as Forms 67-78 – will be administered to Group 1 (Fixed Unit Order) and each form will be taken by 32 students.

44. The FT integrated design with Financial Literacy is shown in Figure 6 with a total sample size of 2,376 students (1,992 students from the original design + 384 students from the FT design).



Forms	Cluster 1	Cluster 2	Cluster 3	Cluster 4
01	S1	S4	M1	M2
02	S3	S6	M3	M4
03	S5	S2	M5	M6ab
04	M2	M3	S2	S5
05	M4	M5	S4	S1
06	M6ab	M1	S6	S3
07	M1	M4	R1	R2
08	M3	M6ab	R3	R4
09	M5	M2	R5	R6ab
10	R2	R3	M2	M5
11	R4	R5	M4	M1
12	R6ab	R1	M6ab	M3
13	R1	R4	S1	S2
14	R3	R6ab	S3	S4
15	R5	R2	S5	S6
16	S2	S3	R2	R5
17	S4	S5	R4	R1
18	S6	S1	R6ab	R3
67	M1	R1	FL1	FL2
68	R2	M2	FL1	FL3
69	M3	R3	FL2	FL1
70	R4	M4	FL2	FL3
71	M5	R5	FL3	FL1
72	R6	M6	FL3	FL2
73	FL1	FL2	M6	R4
74	FL1	FL3	R5	M1
75	FL2	FL1	M4	R2
76	FL2	FL3	R3	M5
77	FL3	FL1	M2	R6
78	FL3	FL2	R1	M3

Forms	Cluster 1	Cluster 2	Cluster 3	Cluster 4
19	R1	R14	R12	R7
20	R2	R16	R14	R9
21	R3	R18	R16	R11
22	R4	R8	R18	R13
23	R5	R10	R8	R15
24	R6ab	R12	R10	R17
25	R13	R1	R10	R9
26	R15	R2	R12	R11
27	R17	R3	R14	R13
28	R7	R4	R16	R15
29	R9	R5	R18	R17
30	R11	R6ab	R8	R7
31	R11	R18	R1	R8
32	R13	R8	R2	R10
33	R15	R10	R3	R12
34	R17	R12	R4	R14
35	R7	R14	R5	R16
36	R9	R16	R6ab	R18
37	R16	R17	R15	R1
38	R18	R7	R17	R2
39	R8	R9	R7	R3
40	R10	R11	R9	R4
41	R12	R13	R11	R5
42	R14	R15	R13	R6ab

Forms	Cluster 1	Cluster 2	Cluster 3	Cluster 4
43	R7	R8	R10	R14
44	R8	R9	R11	R15
45	R9	R10	R12	R16
46	R10	R11	R13	R17
47	R11	R12	R14	R18
48	R12	R13	R15	R7
49	R13	R14	R16	R8
50	R14	R15	R17	R9
51	R15	R16	R18	R10
52	R16	R17	R7	R11
53	R17	R18	R8	R12
54	R18	R7	R9	R13
55	R7	R13	GC1	GC2
56	R8	R14	GC2	GC3
57	R9	R15	GC3	GC4
58	R10	R16	GC4	GC1
59	R11	R17	GC1	GC3
60	R12	R18	GC2	GC4
61	GC1	GC2	R13	R8
62	GC2	GC3	R14	R9
63	GC3	GC4	R15	R10
64	GC4	GC1	R16	R11
65	GC3	GC1	R17	R12
66	GC4	GC2	R18	R7

Figure 6. Field Trial Computer-Based Assessment Design, with Global Competence and Financial Literacy

Where:

- R1-R6ab represent CBA trend Reading clusters
- R7-R18 represent CBA new Reading clusters
- M1-M6ab represent CBA trend Math clusters
- S1-S6 represent CBA trend Science clusters (assembled from 2015 trend and new items)
- GC1-GC4 represent CBA new Global Competence clusters
- FL1-FL3 represent CBA new and trend Financial Literacy clusters

Note: FUO = fixed unit order; VUO = variable unit order

PISA 2018 Questionnaire Design

45. Starting with the first cycle in 2000, PISA has emphasized the importance of collecting background information from students and schools along with the assessment of student achievement. This has been done through a Student Questionnaire of approximately 35 minutes and a School Questionnaire of approximately 45 minutes that covered a broad range of contextual variables. The content of these questionnaires—especially the Student Questionnaire—changed considerably over cycles, but the design remained stable: Every student completed the Student Questionnaire, and every school principal completed the School Questionnaire.

46. PISA has also included several international options, that is, additional instruments that countries could choose to administer. PISA 2018 offers the following optional questionnaires: Educational Career Questionnaire, ICT Familiarity Questionnaire, and Well-Being Student Questionnaire for students. In addition, PISA 2018 also offers a Teacher Questionnaire and a Parent Questionnaire.

47. The background questionnaires contribute to integral aspects of the analytical power of PISA as well as to its capacity for innovation. Therefore, the Questionnaire Design must meet high methodological standards, such as allowing for reliable and unbiased estimation of population parameters for each participating country. In addition, the design also has to ensure that important policy issues and research questions can be addressed in subsequent analysis and reporting based on PISA data. Both the psychometric quality of the variables and indicators and the analytical power of the study have to be taken into account when proposing and evaluating a questionnaire design.

48. The Field Trial and Main Survey questionnaire designs differ in many respects. The main purpose of the Field Trial is to test more material than will be implemented in the Main Survey and to provide the information necessary to determine which scales should be retained for the Main Survey. The proposed designs described below reflect these goals.

49. Annexes A, B and C show the integrated designs, with detailed information about the questionnaires for both the computer-based and paper-based modes respectively as well as the UH computer-based design.

Main Survey Questionnaire Design

50. During the Main Survey, PISA 2018 will implement two compulsory questionnaires for each participating country, a Student Questionnaire and a School Questionnaire.

51. The Student Questionnaire will consist of a single form that will be administered to all students taking the PISA test, with an assessment time of approximately 35 minutes. The questions will cover most of the modules described in the questionnaire framework. For countries choosing to implement international options, the confirmed optional questionnaires are: Educational Career Questionnaire (10 minutes); ICT Familiarity Questionnaire (10 minutes); Well-Being Questionnaire (15 minutes). These optional questionnaires will only be available for countries testing via computer and will be administered to students after the compulsory Student Questionnaire. In addition, for countries that choose to participate in Financial Literacy, there is an 8-10 min FL Questionnaire.

52. Students taking the UH instrument will take the UH version of the Student Questionnaire. This form includes a subset of items from the regular 35-minute Student Questionnaire (mostly trend items)

that will be administered in the same mode as the UH instrument. Also, students taking the UH Booklet-UH Student Questionnaire will not be administered any of the international options.

53. Administrators in each of the 150 participating schools will respond to a School Questionnaire of approximately 45 minutes. Additionally, countries may also choose to implement one or both of the following optional questionnaires: i) a computer-based Teacher Questionnaire and ii) a 30-minute paper-based Parent Questionnaire.

54. Both the Student Questionnaire and the School Questionnaire will be available as paper-based as well computer-based instruments with the mode of the questionnaires matching the country's selected mode of administration of PISA 2018.

55. Finally, the assessment of Financial Literacy will also be accompanied by a short student questionnaire focusing on Financial Literacy education and money matters. So, all students within countries that participate in the optional assessment of Financial Literacy will take this short questionnaire.

Field Trial Questionnaire Design

56. As previously mentioned, one of the major goals of the Field Trial is to evaluate the quality of the context questionnaires used in previous cycles as well as the quality of new items developed for PISA 2018. Thus, more content will be tested in the Field Trial with the goal that some of the new content will be dropped for the Main Survey. Consistent with the Main Survey, PISA 2018 will implement two compulsory questionnaires for each participating country, a student questionnaire and a school questionnaire during the Field Trial.

57. A Student Questionnaire will be administered to all students taking the PISA test, with an assessment time of approximately 35 minutes per student. It will be administered on computer or paper, matching the country's selected mode of administration for PISA 2018.

58. It is important to note that for the Field Trial, two different versions of the Student Questionnaire will be developed to allow for the inclusion of more content. Thus the Field Trial questionnaire will include a set of core items (i.e., StQ-FT Core Items) and one of two rotated blocks (i.e., StQ-FT-A, StQ-FT-B). A minimal set of student background variables—around five minutes in length—will be administered to all students. In contrast, the two rotated blocks will consist of some 30 minutes of non-overlapping content. These two blocks will be randomly assigned to the all students.

59. The confirmed optional questionnaires for students—Educational Career, ICT Familiarity and Well-Being—will be administered following the Student Questionnaire, with a duration of approximately 10-15 minutes each. These are available only as computer-based instruments for countries choosing to implement this option. .

60. The School Questionnaire in the Field Trial will be available in two forms. It will be administered on computer or paper, matching a country's mode of administration of PISA 2018, and will be approximately 45 minutes in length. The slightly longer duration in the Field Trial is because new questions for the School Questionnaire need to be administered and evaluated in order to prepare the final selection of items to be administered in the Main Survey.

61. Additionally, countries may also choose to implement one or both of the following optional questionnaires: i) a computer-based Teacher Questionnaire and ii) a 30-minute paper-based Parent Questionnaire.

62. Finally, similar to the Main Survey, the assessment of Financial Literacy will also be accompanied by a short student questionnaire (8-10 min) focusing on Financial Literacy education and money matters. So, all students within countries that participate in the optional assessment of Financial Literacy will take this short questionnaire.

Annexes

Annex A. Field Trial Computer-Based Integrated Design

Test Session *(120 minutes)* *Computer-Based, Cognitive Test*	
Student Questionnaires *Computer-Based Questionnaires*	
Student Questionnaire (StQ) *(30-35 minutes minutes)* StQ-FT Core Items (5 minutes): *gender, age, grade, educational program, parental occupation, parental education, immigration background* Within-school random assignment to one of two non-overlapping blocks (30 minutes each)	
StQ-FT-A *(30 minutes)*	StQ-FT-B *(30 minutes)*
Optional: ICT Familiarity Questionnaire (ICTQ) *(10 minutes)*	
Optional: Financial Literacy Questionnaire (FLQ) *(8-10 minutes)*	
Optional: Educational Career Questionnaire (ECQ) *(10 minutes)*	
Optional: Well-Being Questionnaire (WBQ) *(10 minutes)*	
School Questionnaire *(45 minutes)* *Computer-Based, Online Questionnaire*	
Teacher Questionnaires (TQ) *(about 60 minutes)* *Computer-Based, Online Questionnaires*	
TQ-FT-Test Language (35 min) To be administered to the sample of test language teachers	TQ-FT-General (35 min) To be administered to the sample of general teachers

Parent Questionnaire (PQ) *(30 minutes)* *Paper-Based Questionnaire*

Annex B. Field Trial Paper-Based Integrated Design

Test Session *(120 minutes)* *Paper-Based, Cognitive Test*

Student Questionnaire *(35 minutes)* *Paper-Based Questionnaire*	
StQ-FT Core Items (5 minutes): gender, age, grade, educational program, parental occupation, parental education, immigration background	
Within-school random assignment to one of two non-overlapping blocks (30 minutes each)	
StQ-FT-A	StQ-FT-B

School Questionnaire *(45 minutes)* *Paper-Based Questionnaire*

Annex C. Field Trial Paper-Based Integrated Design – UH Design

UH Test Session *(60 minutes)* *Computer-Based, Cognitive Test*

UH Student Questionnaire *(20 minutes)* *Computer-Based Questionnaires*