PISA 2022 Technical Report

2 The PISA 2022 Integrated Assessment Design

Introduction

This chapter describes the integrated assessment design for PISA 2022 as well as the processes used by the PISA Core A contractor, Educational Testing Service (ETS) to design the assessment forms for the PISA 2022 cycle.

The cognitive tests for the PISA 2022 cycle included the following:

- a mathematics test, the major domain,
- a reading and a science test, the two minor domains,
- a creative thinking test, the innovative domain, and
- a financial literacy test, an international option.

The development of the mathematics assessment is discussed further in Chapter 3 of this Technical Report. The development of the Creative Thinking domain is presented and discussed in the Chapter 4 of this Technical Report.

PISA 2022 integrated design

The goals for the integrated assessment design in PISA 2022 included:

- continue improving the measurement of trends over time across the three core PISA domains (reading, mathematics, and science),
- continue minimising respondent burden, while maximising the range of information obtained for each domain assessed and from each participating student,
- accurately describing the proficiencies of nationally representative samples of 15-year-olds in each country, including subpopulations of interest, and
- associating these proficiencies with a range of indicators of policy-relevant areas.

To meet these goals, the design for PISA 2022 was based on the design and methodological innovations first introduced in the PISA 2015 cycle and the experience with multistage adaptive testing in the PISA 2018 cycle. In contrast to cycles prior to PISA 2015 where scaling was focused on the cycle at hand and required a new scoring transformation each time, the methodology introduced in PISA 2015 incorporated all then available data for scaling and provided a scoring transformation applicable to PISA 2015 as well as future cycles. It provided a more solid basis for linking across cycles and between paper- and computer-based administrations for all cognitive domains and facilitated the development and transition to computer-based adaptive testing.

As a form of adaptive testing particularly well suited for PISA, multistage adaptive testing was adopted in PISA 2018 for the reading literacy domain. This was adopted with the goal to reduce measurement error across heterogeneous populations without overburdening individual respondents. The experience of the 2018 MSAT and taking note of the differences between reading and mathematics allowed further enhancement of the MSAT design for the mathematics CBA assessment in PISA 2022. Taken together, these design and methodological innovations served to improve comparability across countries/economies, improve parameter estimations and the measurement of trends and improve the reliability of inferences made from the data. In addition, as part of the design for PISA 2022, ETS integrated the domain of creative thinking into the assessment design together with the core domains of reading, mathematics, and science.

Minimising the distinction between major and minor domain coverage

Prior to PISA 2015, the PISA test design focused on keeping the number of students who responded to each item in both the major and minor domains relatively constant. As a result, as shown in Table 2.1, the number of items included in the minor domains was significantly lower than the number of items in the major domain (shown in red font for each cycle). Note, for example, that when mathematics was a minor domain in 2000, 2006, and 2009, it contained about 50% of the items used when it was the major domain in 2003, and between 32-44% when it was the major domain in 2012. Furthermore, when reading was a minor domain in 2003 and 2006, it contained only about 20% of the items used when it was the major domain in 2000.

In contrast, under the assessment design for PISA 2022, 197 items were used in the minor domain of reading, which is 80% of the items when reading was last the major domain in 2018 — and there were 115 items in science, which is 63% of the items when it was last the major domain in 2015. Furthermore, the total number of items across the three core domains increased in ten years from 206 in 2012 to 546 in 2022, an increase of 165%.

Altogether, the inclusion of a larger number of items in each minor domain helped to stabilize and improve the measurement of trend by making the construct coverage for each minor domain more comparable to that of a major domain. The target sample size was not increased accordingly, so there was a reduction of the number of student responses per item for the minor domains. However, since trend items are used for minor domains, there typically is sufficient data for each item by combining the information from the current PISA cycle with that from when the subject was a major domain.

Under this approach for measuring trends, each domain goes through a “domain rotation” over four PISA cycles, that begins with a new or revised framework and continues with the two subsequent cycles in which it becomes a minor domain. The rotation concludes, and starts again, with becoming a major domain on the fourth cycle. The end of the full domain cycle involves a revision of the framework to reflect the current thinking about assessment for the new data collection as a major domain. For example, the revised framework for mathematics as the major domain in PISA 2022 and the introduction of computer-based items broadened the construct beyond what was measured in PISA 2012, the last time that mathematics was a major domain. The framework and instruments for mathematics are expected to remain constant for the next two PISA cycles, with the next revision of the mathematics assessment expected for PISA 2033 when mathematics will again be the major domain.

Multistage adaptive testing

The PISA Governing Board’s (PGB) long-term development strategy for PISA includes the objective of continuing to exploit the advantages of computer-based testing, including the increased use of adaptive testing to further improve measurement accuracy and efficiency, especially at the extremes of the proficiency scale. Additionally, by allowing measurement across a broader range of the ability distribution, adaptive testing could be viewed as making it possible to better measure a more diverse set of participants, thereby extending the global reach of the PISA assessment.

Multistage adaptive testing (MSAT) was introduced in PISA 2018 for the reading major domain only. In PISA 2022, MSAT was extended to the major domain of mathematics, while a reduced MSAT design was created for the now minor domain of reading. The PISA science assessment does not yet follow an adaptive design and one is foreseen to be implemented in PISA 2025. To prepare the MSAT design for mathematics, during the PISA 2022 field trial, unit order was varied to examine whether the order in which units are presented has any impact on item parameter and proficiency estimation. The results of this study in the field trial showed that unit order did not have a significant impact on item parameters nor on proficiency estimates, supporting the use of an MSAT design for mathematics in the main survey. More information about this aspect is provided under the main survey design section of this chapter.

Goals and domain coverage

The design for the PISA 2022 core assessment was developed to provide participating countries/economies with the following information:

- population proficiency distributions in mathematics, the major domain, that reflect the new PISA 2022 mathematics framework and is linked through trend materials to the framework and scale developed in PISA 2012,
- population proficiency distributions in mathematics process and content subscales,
- population proficiency distributions in the minor domain of reading, linked to the PISA 2018 reading framework through trend items for reading,
- population proficiency distributions in the minor domain of science, linked to the PISA 2015 science framework through trend items for science,
- population proficiency distributions in creative thinking, the innovative domain in PISA 2022,
- correlations among the core domains (mathematics, reading, and science) and the innovative domain (creative thinking),
- correlations between mathematics process and content subscales and the other core domains (reading, and science),
- data to link the two modes of delivery: paper- based and computer-based[1].

In addition to the three core domains and the innovative domain, the PISA 2022 assessment also included an optional assessment of financial literacy, which was administered only as a computer-based assessment. For countries/economies participating in the optional domain of financial literacy, population distributions linked to the PISA 2018 financial literacy framework through trend items were provided as well as correlations between financial literacy and mathematics and reading domains.

Overview of the field trial assessment design

The PISA 2022 field trial was designed to provide the information needed in preparation for the main survey. Due to the Covid-19 pandemic, many countries/economies had difficulties with either planning, executing, or completing the data collection for the field trial (see the Field Trial section later in this chapter).

As with the PISA 2018 field trial, the PISA 2022 field trial was designed to verify trend and new items and the feasibility of the integrated design planned for the main survey. In particular, it was designed to verify the feasibility of the new MSAT design for mathematics planned for the main survey and the reduced MSAT design for reading. To ensure appropriate sampling of content, scaling of items and, improved adaptation to student proficiency, the PISA MSAT design offers many alternative options for the selection and delivery of many pre-assembled testlets (i.e., a set of items containing several units) of varying difficulty. As part of the design, units need to be assigned to more than one testlet in different test positions. Thus, while the order of items within a unit does not change, the position of a unit across testlets can be different. For example, a certain unit can be presented as the first unit in some testlets, but as the second unit in others. Therefore, it is important to verify that the psychometric properties of the items and units are invariant when used in different positions (i.e., absence of item/unit position effects). Furthermore, the same unit can be surrounded by different units in different testlets across stages of the MSAT, so that testlets of different difficulty levels are created while ensuring links between them.

The observation of order effects in early PISA cycles (prior to 2015) had led to the assumption that intact cluster positions were needed for parameter invariance to hold. However, 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 (von Davier et al., 2019[1]). The PISA 2022 field trial was designed to provide additional information regarding item parameter invariance under variable unit positions. To that effect, the field trial collected data to study unit order effects by manipulating fixed and variable positions within 30-minute (intact) clusters, and students were randomly assigned to three groups with different unit orders.

For the PISA 2022 field trial, a unit was again considered to represent the minimum granular size of item sets at which adaptiveness can take place. Units consist of a set of items based on a common stimulus or stimuli that can be considered as the organizing grain size that can be assigned randomly or guided by adaptiveness. Although within-unit adaptiveness would be possible in principle, no variations were 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 was a necessary condition for the PISA 2022 mathematics assessment to be delivered in adaptive mode.

With this in mind, the goals of the field trial design included:

- evaluation of the invariance of item parameters compared to previous PISA cycles (both CBA and PBA),
- evaluation of the invariance of item parameters regarding the positions of intact units; that is, a comparison of stability of item parameters between 30-minute clusters found in prior PISA cycles versus varying positions of smaller collections of units to examine the feasibility of introducing MSAT for mathematics in the main survey,
- obtaining preliminary item parameters for the evaluation of new mathematics, financial literacy, and creative thinking 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,
- assessing how well the computer platform functions within and across participating countries/economies.

Like the main survey design, the field trial design for PISA 2022 implemented one CBA design including mathematics, reading, science as core domains, creative thinking as innovative domain, and financial literacy as the optional domain. In addition, the field trial design also included two PBA designs that involved the three core domains of mathematics, reading, and science. One PBA design was the same as implemented in PISA 2018. The other, new PBA design was developed for newly participating countries/economies. The new PBA instrument was the same one that was used for PISA for Development[2].

The standard design for countries/economies choosing computer delivery for the assessment was to select a minimum of 28 schools for the field trial and select 71-72 students within each school. This design resulted in a sample size of approximately 2,000 assessed students. Alternative designs to achieve the same sample size were available for participants having difficulty in finding enough large schools where to implement this design.

Countries/economies that chose to participate using only paper-and-pencil forms had a reduced sample-size requirement. The goals for these participants were mainly focused on testing operations and data-processing related procedures. For both the PBA and new PBA designs, these participants selected 25 schools with 36 students from each school for a total field trial sample of approximately 900 assessed students.

Field trial CBA design

The computer-based assessment (CBA) design for the field trial organized the items into 69 different test forms and students into three groups. Students in groups 1 and 3 took fixed-unit order (FUO) forms, while students in group 2 took variable-unit order (VUO) forms. The standard field trial CBA design is shown in Figure 2.1. Each test form consisted of 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. Each cluster consisted of multiple units, and the ordering of the units was always fixed and consistent in FUO forms. In contrast, ordering of the units was varied across VUO forms. For example, cluster M1 cluster in form 19 had a different ordering of units compared to the ordering of units in cluster M1 in form 25. More specifically, students in group 1 took forms 1–18 with trend items in mathematics, reading, and science. Group 2 took 24 forms (forms 19-42) with both new and trend mathematics items. Group 3 took 27 forms with either only new mathematics items (forms 43–54) or new mathematics and creative thinking items (forms 55-69). Students in group 1 who took reading were administered the reduced MSAT design discussed later in this chapter. Furthermore, the same set of 65 sentences from the 2018 Main Survey were used to measure reading fluency as part of the Reading scale.

Field trial PBA designs

As noted, there were two PBA instruments offered this PISA cycle. The first PBA design was a version administered by only one participant and contained the same trend clusters that were administered in PISA 2015 and PISA 2018 for paper-based participants. The second PBA was new for this PISA cycle. However, the materials have previously been administered in PISA for Development and were successfully linked to the PISA scales as there are items common to both instruments. This new PBA instrument was administered by all other PBA participants. Under the first PBA design, students were randomly assigned one of the 18 PBA forms that contained trend items from two of the three core domains for PISA – reading, mathematics, and science. This design is shown in Figure 2.2.

Students in countries/economies that chose the second, new PBA design were randomly assigned one of 12 new PBA forms that contained trend items from two of the three core domains for PISA – mathematics, science, and reading/reading fluency. This design is also shown in Figure 2.2.

Overview of the main survey assessment design

The assessment design for PISA 2022 was planned so that the total testing time was two hours for each student, followed by a student background questionnaire. An overview of the flow of the integrated design for the PISA 2022 main survey is presented in Figure 2.3.

Figure 2.1. Field trial computer-based assessment design

	Forms	Cluster 1	Cluster 2	Cluster 3	Cluster 4
GROUP 1 CBA Trend FUO (Forms 01-18)	1	S1	S4	M1	M2
	2	S3	S6	M3	M4
	3	S5	S2	M5	M6ab
	4	M2	M3	S2	S5
	5	M4	M5	S4	S1
	6	M6ab	M1	S6	S3
	7	M1	M4	R adaptive	R adaptive
	8	M3	M6ab	R adaptive	R adaptive
	9	M5	M2	R adaptive	R adaptive
	10	R adaptive	R adaptive	M2	M5
	11	R adaptive	R adaptive	M4	M1
	12	R adaptive	R adaptive	M6ab	M3
	13	R adaptive	R adaptive	S1	S2
	14	R adaptive	R adaptive	S3	S4
	15	R adaptive	R adaptive	S5	S6
	16	S2	S3	R adaptive	R adaptive
	17	S4	S5	R adaptive	R adaptive
	18	S6	S1	R adaptive	R adaptive
GROUP 2 CBA Trend M/New M VUO (Forms 19-42)	19	M1	M14	M12	M7
	20	M2	M16	M14	M9
	21	M3	M18	M16	M11
	22	M4	M8	M18	M13
	23	M5	M10	M8	M15
	24	M6ab	M12	M10	M17
	25	M13	M1	M10	M9
	26	M15	M2	M12	M11
	27	M17	M3	M14	M13
	28	M7	M4	M16	M15
	29	M9	M5	M18	M17
	30	M11	M6ab	M8	M7
	31	M11	M18	M1	M8
	32	M13	M8	M2	M10
	33	M15	M10	M3	M12
	34	M17	M12	M4	M14
	35	M7	M14	M5	M16
	36	M9	M16	M6ab	M18
	37	M16	M17	M15	M1
	38	M18	M7	M17	M2
	39	M8	M9	M7	M3
	40	M10	M11	M9	M4
	41	M12	M13	M11	M5
	42	M14	M15	M13	M6ab
GROUP 3 CBA New M/CRT FUO (Forms 43-69)	43	M7	M8	M10	M14
	44	M8	M9	M11	M15
	45	M9	M10	M12	M16
	46	M10	M11	M13	M17
	47	M11	M12	M14	M18
	48	M12	M13	M15	M7
	49	M13	M14	M16	M8
	50	M14	M15	M17	M9
	51	M15	M16	M18	M10
	52	M16	M17	M7	M11
	53	M17	M18	M8	M12
	54	M18	M7	M9	M13
	55	M7	M13	CT1	CT2
	56	M8	M14	CT2	CT3
	57	M9	M15	CT3	CT4
	58	M10	M16	CT4	CT5
	59	M11	M17	CT5	CT1
	60	CT3	CT5	M14	M9
	61	CT4	CT1	M15	M10
	62	CT5	CT2	M16	M11
	63	CT1	CT3	M17	M12
	64	CT2	CT4	M18	M7
	65	CT1	CT2	CT3	CT5
	66	CT2	CT3	CT4	CT1
	67	CT3	CT4	CT5	CT2
	68	CT4	CT5	CT1	CT3
	69	CT5	CT1	CT2	CT4

FUO = fixed unit order; VUO = variable unit order
Where:
R adaptive represents CBA trend reading units (containing trend and new items from 2018)
M7-M18 represent CBA new mathematics clusters
M1-M6ab represent CBA trend mathematics clusters (in the 2022 FT, all CBA participants administered both M6a and M6b)
S1-S6 represent CBA trend science clusters (containing trend and new items from 2015)
CT1-CT5 represent CBA new creative thinking clusters

Figure 2.2. Field trial paper-based assessment designs

Design 1 - PBA Design

		Booklets	Cluster 1	Cluster 2	Cluster 3	Cluster 4
900 assessed students (25 schools, 36 students per school)	P=0.47	1	PS1	PS4	PM1	PM2
		2	PS3	PS6	PM3	PM4
		3	PS5	PS2	PM5	PM6b
		4	PM2	PM3	PS2	PS5
		5	PM4	PM5	PS4	PS1
		6	PM6b	PM1	PS6	PS3
	P=0.47	7	PM1	PM4	PR1	PR2
		8	PM3	PM6b	PR3	PR4
		9	PM5	PM2	PR5	PR6b
		10	PR2	PR3	PM2	PM5
		11	PR4	PR5	PM4	PM1
		12	PR6b	PR1	PM6b	PM3
	P=0.06	13	PR1	PR4	PS1	PS2
		14	PR3	PR6b	PS3	PS4
		15	PR5	PR2	PS5	PS6
		16	PS2	PS3	PR2	PR5
		17	PS4	PS5	PR4	PR1
		18	PS6	PS1	PR6b	PR3

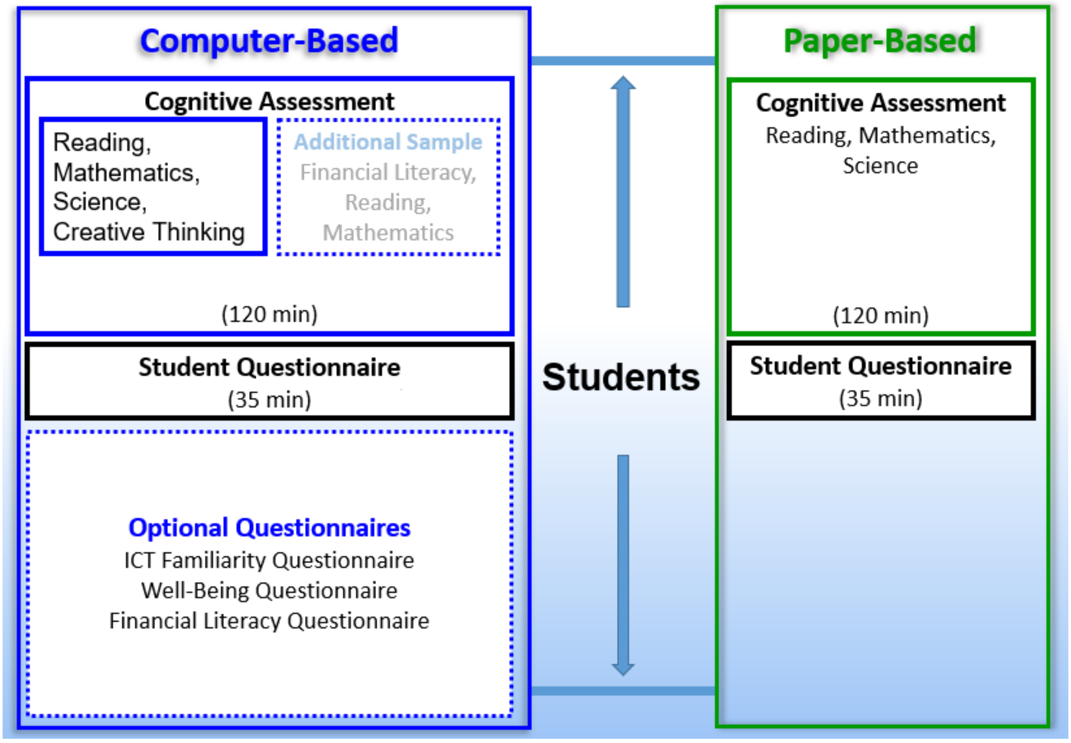
Where:
PR1-PR6b represent PBA trend reading clusters (the participant only administered R6b) - same clusters from 2015 and 2018
PM1-PM6b represent PBA trend mathematics clusters (the participant only administered M6b) - same clusters from 2015 and 2018
PS1-PS6 represent PBA trend science clusters (same clusters from 2015 and 2018)

Design 2 - "new" PBA design

		Booklets	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
900 assessed students (25 schools, 36 students per school)	P=1.00	1	RC1	R1	R2	S1	S2
		2	S2	S3	RC2	R2	R3
		3	RC3	R3	R4	S3	S4
		4	S4	S1	RC4	R4	R1
		5	S1	S2	M1	M2	
		6	M2	M3	S2	S3	
		7	S3	S4	M3	M4	
		8	M4	M1	S4	S1	
		9	M1	M2	RC1	R1	R2
		10	RC2	R2	R3	M2	M3
		11	M3	M4	RC3	R3	R4
		12	RC4	R4	R1	M4	M1

Where:
RC1-RC4 represent reading components clusters
R1-R4 represent new PBA reading clusters
M1-M4 represent new PBA mathematics clusters
S1-S4 represent new PBA science clusters
Booklets 5-8 did not contain a reading components cluster

Figure 2.3. Overview of the PISA 2022 main survey integrated design



Paper-based integrated designs

For the participant in the first PBA design, the main survey included the same 18 forms as in the field trial assessment design, but sample size requirements differed. The main survey PBA design is shown in Figure 2.4. The PBA test forms did not include any newly developed items. Each form included two of the three core domains with two 30-minute clusters for each domain assessed. As a result, all students were administered four clusters, 47% of participating students were administered two clusters of science items and two clusters of mathematics items, 47% were administered two clusters of mathematics and two clusters of reading, and 6% were administered two clusters of reading and two clusters of science. The PBA was to be administered to 35 students in each of 150 schools, resulting in a total sample size of 5,250 assessed students.

The main survey assessment design for countries/economies that chose the new PBA design included 12 forms (see Figure 2.4) and was the same as for the field trial. These PBA test forms consisted of existing items from PISA for Development. Each form included two of the three core domains with two 30-minute clusters for each domain assessed. Students were administered a randomly selected form. As a result, 33% of participating students were administered two clusters of reading items and two clusters of science items, 33% were administered two clusters of science and two clusters of mathematics, and 33% were administered two clusters of mathematics and two clusters of reading. As with the first PBA design, the new PBA design was to be administered to 35 students in each of 150 schools, resulting in a total of 5,250 assessed students.

Figure 2.4. Main survey paper-based assessment designs

The field trial and main survey paper-based assessment designs were the same with respect to the items/units and clusters, number of booklets, and the order of the clusters within the booklets.

Design 1 - PBA Design

		Booklets	Cluster 1	Cluster 2	Cluster 3	Cluster 4
5,250 assessed students (150 schools, 35 students per school)	P=0.47	1	PS1	PS4	PM1	PM2
		2	PS3	PS6	PM3	PM4
		3	PS5	PS2	PM5	PM6b
		4	PM2	PM3	PS2	PS5
		5	PM4	PM5	PS4	PS1
		6	PM6b	PM1	PS6	PS3
	P=0.47	7	PM1	PM4	PR1	PR2
		8	PM3	PM6b	PR3	PR4
		9	PM5	PM2	PR5	PR6b
		10	PR2	PR3	PM2	PM5
		11	PR4	PR5	PM4	PM1
		12	PR6b	PR1	PM6b	PM3
	P=0.06	13	PR1	PR4	PS1	PS2
		14	PR3	PR6b	PS3	PS4
		15	PR5	PR2	PS5	PS6
		16	PS2	PS3	PR2	PR5
		17	PS4	PS5	PR4	PR1
		18	PS6	PS1	PR6b	PR3

Design 2 - new PBA Design

		Booklets	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
5,250 assessed students (150 schools, 35 students per school)	P=0.33	1	RC1	R1	R2	S1	S2
		2	S2	S3	RC2	R2	R3
		3	RC3	R3	R4	S3	S4
		4	S4	S1	RC4	R4	R1
	P=0.33	5	S1	S2	M1	M2	
		6	M2	M3	S2	S3	
		7	S3	S4	M3	M4	
		8	M4	M1	S4	S1	
	P=0.33	9	M1	M2	RC1	R1	R2
		10	RC2	R2	R3	M2	M3
		11	M3	M4	RC3	R3	R4
		12	RC4	R4	R1	M4	M1

Where:
PR1-PR6b represent PBA trend reading clusters (the participant only administered R6b) - same clusters from 2015 and 2018
PM1-PM6b represent PBA trend mathematics clusters (the participant only administered M6b) - same clusters from 2015 and 2018
PS1-PS6 represent PBA trend science clusters (same clusters from 2015 and 2018)

Where:
RC1-RC4 represent reading components clusters
R1-R4 represent new PBA reading clusters
M1-M4 represent new PBA mathematics clusters
S1-S4 represent new PBA science clusters
Booklets 5-8 did not contain a reading components cluster

Computer-based integrated design

For CBA participants that also administered the creative thinking assessment, the main survey included 66 forms (forms 01-66) which are shown in Figure 2.5. Under the full integrated design that included all four domains, 94% of the sampled students responded to 60 minutes of mathematics items, 39% responded to reading items, 39% to science items, and 28% responded to creative thinking items. As in PISA 2018, sixty-five reading fluency items assigned to six blocks were used. Each student taking reading received two blocks of sentences which were rotated as shown in Figure 2.5.

For countries/economies not participating in the creative thinking assessment, only 36 forms were included in the design (forms 01-36). The percentages for this alternative design are also represented in Figure 2.6.

Main survey multistage adaptive testing design: Mathematics and Reading

The MSAT design that was implemented for mathematics in the PISA 2022 main survey was built upon the MSAT design used for reading in PISA 2018. However, using the experience from PISA 2018 and the differences between mathematics and reading, it was possible to enhance the following four areas:

1. *Balancing the MSAT design.* A fully balanced design was implemented so that each item occurred in every stage, this to further address potential position effects. This feature is similar to the balanced incomplete block (BIB) designs used in previous, non-adaptive PISA cycles.
2. *More adaptivity.* A third level of difficulty was introduced in the third stage, which was possible because there were more machine-scored items and smaller units in mathematics than there were in reading.
3. *Linear component.* A hybrid design with an adaptive and linear component was used so that the probability layer used in the PISA 2018 MSAT design for reading could be eliminated. The probability layer used determined the difficulty of the next set of items to be administered, with a low probability assigned to a misrouting. Instead of this probability layer, 25% of students were administered a linear test to avoid the intentional misrouting of students to items that would be either too easy or too difficult for them).
4. *Automated assembly.* Formal methods for optimal design and test assembly were employed by making use of linear programming techniques, which provided a principled approach to support the decision-making process for the MSAT design.

Since reading was not the major domain this cycle, the MSAT reading design used for PISA 2022 was a reduced version of the MSAT design used in PISA 2018. That is, the same number of stages and adaptive levels were used, but with a smaller item pool (about 25% fewer items, 196 instead of 245 items) and fewer testlets (30 instead of 40 testlets). As in PISA 2018, each student assessed in reading received seven units. In design A (75%), students take 2, 3, and 2 reading units across the three stages from three sets of units, whereas students take 2, 2, and 3 reading units, respectively, in design B (25%) where the unit sets for the last two stages are reversed compared to design A. The same probability layer was used as in PISA 2018 for routing students through different MSAT paths (see PISA 2018 Tech Report, Chapter 2). In PISA 2022, each student assessed in reading responded to 35-42 reading items, while in PISA 2018 the range was 33-40 items. The PISA 2022 design still allowed students to take approximately the same number of items within the same amount of assessment time.

Figure 2.5. Main survey computer-based assessment design

Percentage of Students	Forms	Fluency	Cluster 1	Cluster 2	Fluency	Cluster 3	Cluster 4
35% (No CT= 48%)	1		$M_{(adaptive)}$		fl1	$R_{(adaptive)}$	
	2		$M_{(adaptive)}$		fl2	$R_{(adaptive)}$	
	3		$M_{(adaptive)}$		fl3	$R_{(adaptive)}$	
	4		$M_{(adaptive)}$		fl4	$R_{(adaptive)}$	
	5		$M_{(adaptive)}$		fl5	$R_{(adaptive)}$	
	6		$M_{(adaptive)}$		fl6	$R_{(adaptive)}$	
	7	fl7	$R_{(adaptive)}$			$M_{(adaptive)}$	
	8	fl8	$R_{(adaptive)}$			$M_{(adaptive)}$	
	9	fl9	$R_{(adaptive)}$			$M_{(adaptive)}$	
	10	fl10	$R_{(adaptive)}$			$M_{(adaptive)}$	
	11	fl11	$R_{(adaptive)}$			$M_{(adaptive)}$	
	12	fl12	$R_{(adaptive)}$			$M_{(adaptive)}$	
35% (No CT= 48%)	13		$M_{(adaptive)}$			S1	S2
	14		$M_{(adaptive)}$			S2	S3
	15		$M_{(adaptive)}$			S3	S4
	16		$M_{(adaptive)}$			S4	S5
	17		$M_{(adaptive)}$			S5	S6
	18		$M_{(adaptive)}$			S6	S1
	19		S1	S3		$M_{(adaptive)}$	
	20		S2	S4		$M_{(adaptive)}$	
	21		S3	S5		$M_{(adaptive)}$	
	22		S4	S6		$M_{(adaptive)}$	
	23		S5	S1		$M_{(adaptive)}$	
	24		S6	S2		$M_{(adaptive)}$	
2% (No CT= 4%)	25	fl1	$R_{(adaptive)}$			S1	S2
	26	fl2	$R_{(adaptive)}$			S2	S3
	27	fl3	$R_{(adaptive)}$			S3	S4
	28	fl4	$R_{(adaptive)}$			S4	S5
	29	fl5	$R_{(adaptive)}$			S5	S6
	30	fl6	$R_{(adaptive)}$			S6	S1
	31		S1	S3	fl7	$R_{(adaptive)}$	
	32		S2	S4	fl8	$R_{(adaptive)}$	
	33		S3	S5	fl9	$R_{(adaptive)}$	
	34		S4	S6	fl10	$R_{(adaptive)}$	
	35		S5	S1	fl11	$R_{(adaptive)}$	
	36		S6	S2	fl12	$R_{(adaptive)}$	
24% (No CT= NA)	37		$M_{(adaptive)}$			CT1	CT2
	38		$M_{(adaptive)}$			CT2	CT3
	39		$M_{(adaptive)}$			CT3	CT4
	40		$M_{(adaptive)}$			CT4	CT5
	41		$M_{(adaptive)}$			CT5	CT1
	42		CT2	CT4		$M_{(adaptive)}$	
	43		CT3	CT5		$M_{(adaptive)}$	
	44		CT4	CT1		$M_{(adaptive)}$	
	45		CT5	CT2		$M_{(adaptive)}$	
	46		CT1	CT3		$M_{(adaptive)}$	
2% (No CT= NA)	47	fl1	$R_{(adaptive)}$			CT1	CT2
	48	fl2	$R_{(adaptive)}$			CT2	CT3
	49	fl3	$R_{(adaptive)}$			CT3	CT4
	50	fl4	$R_{(adaptive)}$			CT4	CT5
	51	fl5	$R_{(adaptive)}$			CT5	CT1
	52		CT2	CT4	fl7	$R_{(adaptive)}$	
	53		CT3	CT5	fl8	$R_{(adaptive)}$	
	54		CT4	CT1	fl9	$R_{(adaptive)}$	
	55		CT5	CT2	fl10	$R_{(adaptive)}$	
	56		CT1	CT3	fl11	$R_{(adaptive)}$	
2% (No CT= NA)	57		S1	S3		CT1	CT2
	58		S2	S4		CT2	CT3
	59		S3	S5		CT3	CT4
	60		S4	S6		CT4	CT5
	61		S5	S1		CT5	CT1
	62		CT2	CT4		S1	S2
	63		CT3	CT5		S2	S3
	64		CT4	CT1		S3	S4
	65		CT5	CT2		S4	S5
	66		CT1	CT3		S6	S1

Where:
R(adaptive) represents the computer-based reading assessment (trend) in an adaptive design
M(adaptive) represents the computer-based mathematics assessment (trend and new) in an adaptive design
S1-S6 represent the computer-based science clusters (trend)
CT1-CT5 represent the computer-based creative thinking clusters (new)
fl1-fl12 represent the computer-based reading fluency clusters (trend and new items)

Figure 2.6. Overview of main survey computer-based MSAT design - with creative thinking and without creative thinking

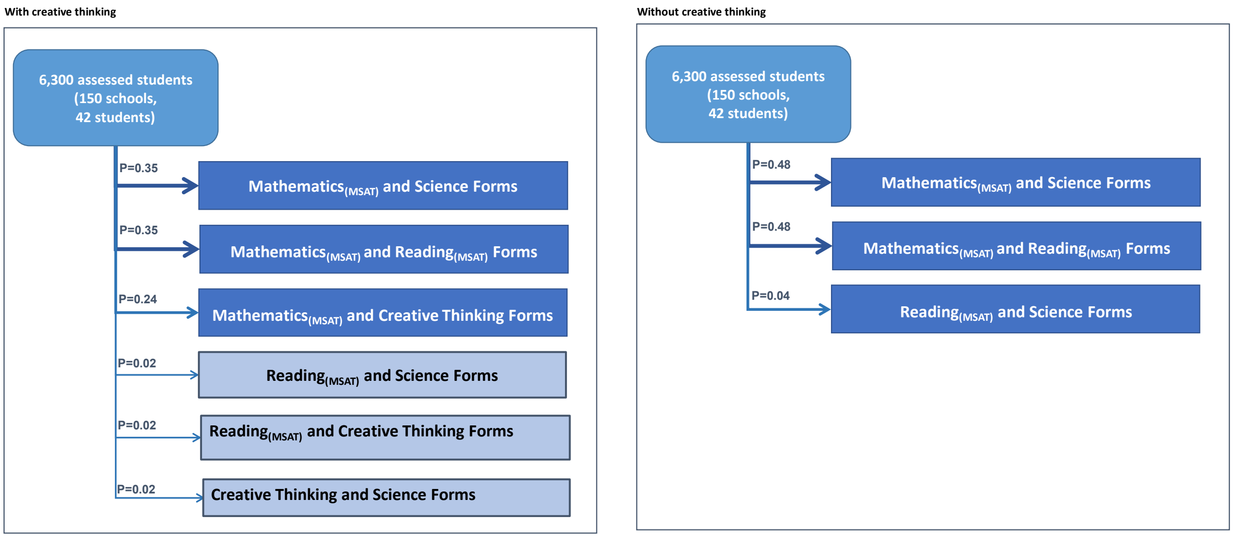


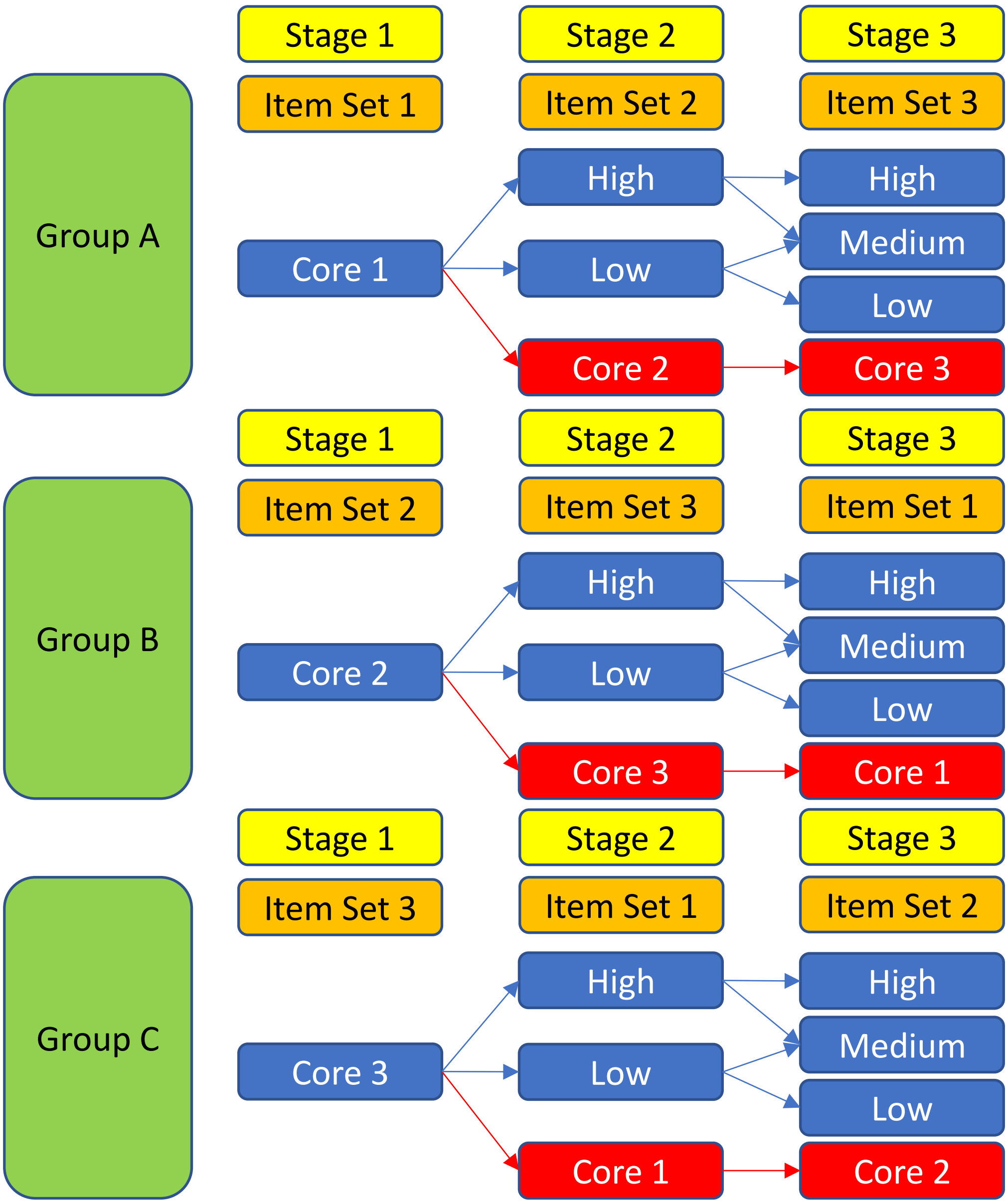
Figure 2.7 shows an overview of the hybrid MSAT design used for mathematics in the PISA 2022 main survey. The MSAT design for mathematics consisted of three stages and 234 mathematics items from a total of 99 units. The items were divided into three equivalent and mutually exclusive item sets, each consisting of 78 items from 33 units. From each item set, 16 testlets of nine or 10 items were created within each stage, so across the three item sets and three stages, there was a total 16*3*3 = 144 testlets. Each student took one testlet in each stage, so the total number of mathematics items taken by each student ranged from 28-30. Within-stage linking was accomplished by having each item appear two, or sometimes three, times across testlets associated with each stage and each group (but no more than seven times overall). For students taking the adaptive part of the design, stage 1 consisted of a core testlet of medium difficulty, stage 2 consisted of high- or low-difficulty testlets, and stage 3 consisted of high-, medium-, or low-difficulty testlets administered in a rotating order to constitute three sets of equivalent instruments that were assigned to three groups of randomly selected students (A, B, and C). For students that were assigned to the linear part of the design, after the stage 1 core testlet, they proceeded to take a core testlet from the other item sets at each subsequent stage. Figure 2.8 shows the testlet structure for one group (Group A) and the item set associated with that group, as well as four example paths that a student could take under the adaptive part of the design.

The total number of paths in the hybrid MSAT design for mathematics was 240 (see Table 2.2). For the adaptive component, there were 192 total paths since every testlet in stage 1 was associated with four possible paths (going from Stage 1 > Stage 2 > Stage 3):

1. Core > Low > Low
2. Core > Low > Medium
3. Core > High > Medium
4. Core > High > High

For the linear component, a simplified design was chosen where each testlet in stage 1 was associated with one fixed path that resulted in 48 linear forms. The forms are shown in Table 2.3.

Figure 2.7. Overview of the hybrid main survey computer-based MSAT design for mathematics



Where:
Groups A, B, and C represent groups of randomly selected students
Blue represents adaptive parts - taken by 75% of students
Red represents linear parts - taken by 25% of students

Figure 2.8. Example testlet structure across stages for one group

Where:
A represents one group of randomly selected students. The structure is the same for groups B and C, and for the item sets associated with each stage for those groups.
The arrows represent an example of four possible paths. By design, some combinations of testlets were not allowed.

The difficulty of the testlets was targeted by using subsets of the item pool as the statistical target. The average difficulty in stage 1 was targeted by using 100% of the items. At stage 2, low difficulty testlets were targeted by using 75% of the easiest items, and high difficulty levels were targeted by using 75% of the hardest items. At stage 3, a similar approach was taken for low, medium, and high difficulty levels by using 50% of the easiest items, 50% of medium difficulty items, and 50% of the hardest items.

Technically, this targeting was accomplished by using the test information function (TIF) of the relevant subsets of items as the statistical target in the assembly. However, since differences in difficulty can still arise when only the TIF is used [see, e.g., Ali and van Rijn (2016[2])], constraints on the test characteristic curve (TCC) were used as well. The method resulted in the high difficulty testlets at stage 3 being more difficult than the high difficulty testlets at stage 2, and the low difficulty testlets at stage 3 were less difficult than the low difficulty testlets at stage 2, which is ideal because more is known about a student's mathematical proficiency after two stages of assessment.

Additionally, to avoid students experiencing a large shift in difficulty levels between stages, as well as to keep the number of possible paths to a more reasonable number, students who received a low difficulty testlet in stage 2 could not be routed to a high difficulty testlet in stage 3, and students who received a high difficulty testlet in stage 2 could not be routed to a low difficulty testlet in stage 3.The effect of restricting the possible paths is minimal because there is a considerable amount of overlap in the difficulty ranges of testlets of adjacent difficulty (i.e., low/medium and medium/high).

Cut-off values for determining how to route students were identified by first computing the intersections of the average information functions of the testlets. On the PISA mathematics scale, the intersection of low and high difficulty testlets at stage 2 was found to be 495. At stage 3, the intersection between low and medium was found to be 425, and between medium and high was 550. Once these values were identified, the inverse TCC was used to determine the cut scores based on the items within each testlet that could be automatically scored. The cut scores were used to determine a student's path as each stage was completed. Simulation studies showed that this approach would result in about one third of students being routed to each of the difficulty levels at stage 3 for a country/economy that performs around 500 – the midpoint of the scale.

Annex 2.Aof this chapter provides a list of the cut scores, including the maximum score from machine-coded items and the maximum possible score, for each core testlet. Annex 2.Bof this chapter shows cut scores for each adaptive path, including the maximum score from machine-coded items and the maximum possible score. These cut scores are based on the number of raw score points obtained on the machine scored items alone.

Une Heure (UH) form

Consistent with previous cycles, a special one-hour test, referred to as the "Une Heure" (UH) form, was prepared for students with special needs. The selected items were among the easier trend items (i.e., items developed prior to PISA 2015) in each core domain and had a reduced reading load. The UH form contained about half as many items as the other forms, with each cluster including from seven to nine items. In PISA 2022 the UH form was comprised of about 53% mathematics, 21% reading, and 26% science items.

The UH form included two 15-minute clusters of mathematics (MU1 and MU2), one 15-minute cluster of reading (RU1) and one 15-minute cluster of science (SU1). The assignment of this form followed the approach described previously for the assignment of the base test form. The UH form was assigned base form 99 (as shown in Figure 2.9).

Figure 2.9. Main survey computer-based UH form design

Form	Cluster 1	Cluster 2	Cluster 3	Cluster 4
99 (UH)	MU1	MU2	RU1	SU1

Where M = mathematics, R = reading, and S = science

The UH form was accompanied by a special UH student background questionnaire that included only a subset of items from the regular background questionnaire (primarily trend items) in a single form design that was administered in CBA only. No PBA participants chose to administer the UH Form.

Assessment of financial literacy

The assessment of financial literacy was again offered as an international option in PISA 2022. The cognitive instruments included trend items from the PISA 2012, PISA 2015, and PISA 2018 assessments, plus a few new units that were developed for PISA 2022. Financial literacy was administered only as a computer-based assessment to an additional sample of students at the same schools sampled for PISA.

As in PISA 2018, the financial literacy assessment was administered to a separate sample of PISA-eligible students who took, in addition to the financial literacy assessment, a combination of reading or mathematics items. The total testing time for each student was two hours (120 minutes). The sample of students who took the financial literacy assessment are referred to as the “Financial Literacy sample”.

Field trial design for the financial literacy assessment

For the 2022 field trial of the financial literacy assessment, the main sample was augmented by adding a sample of approximately 253 students who were assigned one of the 12 financial literacy testing forms. These forms included 60 minutes of financial literacy items and either 60 minutes of reading items or 60 minutes of mathematics items. These were based on using two financial literacy clusters (F1 and F2), MSAT reading items, and six of the seven trend mathematics clusters (M1 to M6ab). The design is shown in Figure 2.10. The 12 financial literacy forms were administered to Group 1 students (FUO) and each form was administered to about 32 students within each country/economy.

Figure 2.10. Field trial computer-based financial literacy design

Form		Cluster 1	Cluster 2		Cluster 3	Cluster 4
70		M1	M2		F1	F2
71		M3	M4		F2	F1
72		M5	M6ab		F1	F2
73	fl1	R(adaptive)			F2	F1
74	fl2	R(adaptive)			F1	F2
75	fl3	R(adaptive)			F2	F1
76		F2	F1		M2	M5
77		F1	F2		M4	M1
78		F2	F1		M6ab	M3
79		F1	F2	fl4	R(adaptive)	
80		F2	F1	fl5	R(adaptive)	
81		F1	F2	fl6	R(adaptive)	

Where:
F1-F2 represent the computer-based financial literacy clusters (new and trend)
R(adaptive) represents the computer-based reading assessment (trend and new) in an adaptive design
M1-M6ab represent the computer-based mathematics trend clusters
fl1-fl6 represent reading fluency clusters

Main survey financial literacy design

For the main survey, countries/economies participating in the financial literacy assessment were required to assess 1,650 additional students. Each student that took the financial literacy assessment took 60 minutes of financial literacy items, and then either mathematics or reading items. Students taking the financial literacy assessment did not take any of the science items and therefore they do not have science literacy proficiency estimates.

The main survey version of the assessment instruments included 46 financial literacy items, of which 41 were trend items and 5 were new items. These items were organized into two 30-minute clusters of financial literacy (F1 and F2) that were rotated into eight forms each containing 60 minutes of financial literacy and 60 minutes of either MSAT mathematics or MSAT reading items, as shown in Figure 2.11.

Figure 2.11. Main survey computer-based financial literacy design

Form	Cluster 1	Cluster 2	Fluency	Cluster 3	Cluster 4
67	$M_{(adaptive)}$			F1	F2
68	$M_{(adaptive)}$			F2	F1
69	F1	F2		$M_{(adaptive)}$	
70	F2	F1		$M_{(adaptive)}$	
71	$R_{(adaptive)}$			F1	F2
72	$R_{(adaptive)}$			F2	F1
73	F1	F2	fl7	$R_{(adaptive)}$	
74	F2	F1	fl8	$R_{(adaptive)}$	

Where:
F1-F2 represent the computer-based financial literacy clusters (new and trend)
R(adaptive) represents the computer-based reading assessment (trend and new) in an adaptive design
M(adaptive) represents the computer-based mathematics assessment (trend and new) in an adaptive design
fl7-fl8 represent reading fluency clusters

Assigning mathematics units to the multistage adaptive design

As noted earlier, the MSAT design for mathematics expanded and enhanced what was accomplished with the adaptive design for reading in PISA 2018. Test assembly for PISA 2022 was implemented in four steps:

1. Assemble non-overlapping parallel item sets.
2. Assemble core and adaptive testlets from each item set.
3. Assemble multistage adaptive paths using the core and adaptive testlets.
4. Assemble linear forms using the core testlets.

In each step, mixed-integer linear programming was used (van der Linden, 2005[3]; Diao and van der Linden, 2011[4]; van Rijn et al., 2022[5]). In the first step, the decision variables were defined as which unit was to be in which item set. For the second step, the decision variables were defined as which unit was to be in which testlet. In the third step, they were to describe which of the core and adaptive testlets was in which multistage adaptive path. Finally, in step four, they indicated which core testlets were in which linear form. Furthermore, all steps but the first consisted of multiple assemblies (e.g., in step 2, 16 core testlets were assembled from item set A, 16 core testlets were from item set B, etc.)

The objective in each step was always to minimize the difference with respect to a target TIF. In each step, constraints on the following variables were set: item exposure, number of units, number of items, maximum score, maximum human score, number of trend/new items, number of dichotomous/polytomous items, item format, content subdomain, process subdomain, overlap, median response time, and TCC.

As an example, the assembly of a set of core testlets is illustrated. In this case, the main decision variables of the assembly are defined as follows

$$x_{ut} = \begin{cases} 1, & \text{if unit u in testlet t,} \\ 0, & \text{otherwise.} \end{cases}$$

Under local independence, the information function of a unit is the sum of item information functions : $I_u(\theta) = \sum_{i \in V_u} I_i(\theta)$, where V_u indicates the set of items in unit *u*. Similarly, the unit characteristic curve (i.e., the expected score on a unit as a function of θ) is the sum of item characteristic curves : $T_u(\theta) = \sum_{i \in V_u} T_i(\theta)$. The target TIF is denoted by $\mathcal{I}(\theta)$ and the objective is to minimize ϵ subject to

$$\mathcal{I}(\theta_j) - \epsilon \leq \sum_{u=1}^{U} I_u(\theta_j) x_{ut} \leq \mathcal{I}(\theta_j) + \epsilon, \quad \text{for all } j \text{ and } t,$$

where $\epsilon > 0$ and *U* is the number of units in the used item set. For the core testlets, the target TIF was set proportional to the TIF of the item set. The number of θ points, indexed by *j*, at which to evaluate the TIF was three. To avoid potential differences in the TCC, an interval of one score point around the target TCC, $\mathcal{T}(\theta_j)$, was allowed, which can be formalized as

$$\mathcal{T}(\theta_j) - 0.5 \leq \sum_{u=1}^{U} T_u(\theta_j) x_{ut} \leq \mathcal{T}(\theta_j) + 0.5, \quad \text{for all } j \text{ and } t.$$

Other constraints of category *c* can be formulated as:

$$n_c^{\min} \leq \sum_{u=1}^{U} n_{cu}\, x_{ut} \leq n_c^{\max}, \quad \text{for all } t,$$

where $n_c^{\min}$ is the minimum required number (e.g., the number of items, the maximum score), n_{cu} is the number for category *c* of unit *u*, and $n_c^{\max}$ is the maximum required number. Note that the constraints here can be both categorical and numerical. For the core testlets, the number of items was constrained to either 9 or 10 and the maximum score to 12 or 13. Bounds on the number of common items between testlets (overlap) can be added with the following set of constraints:

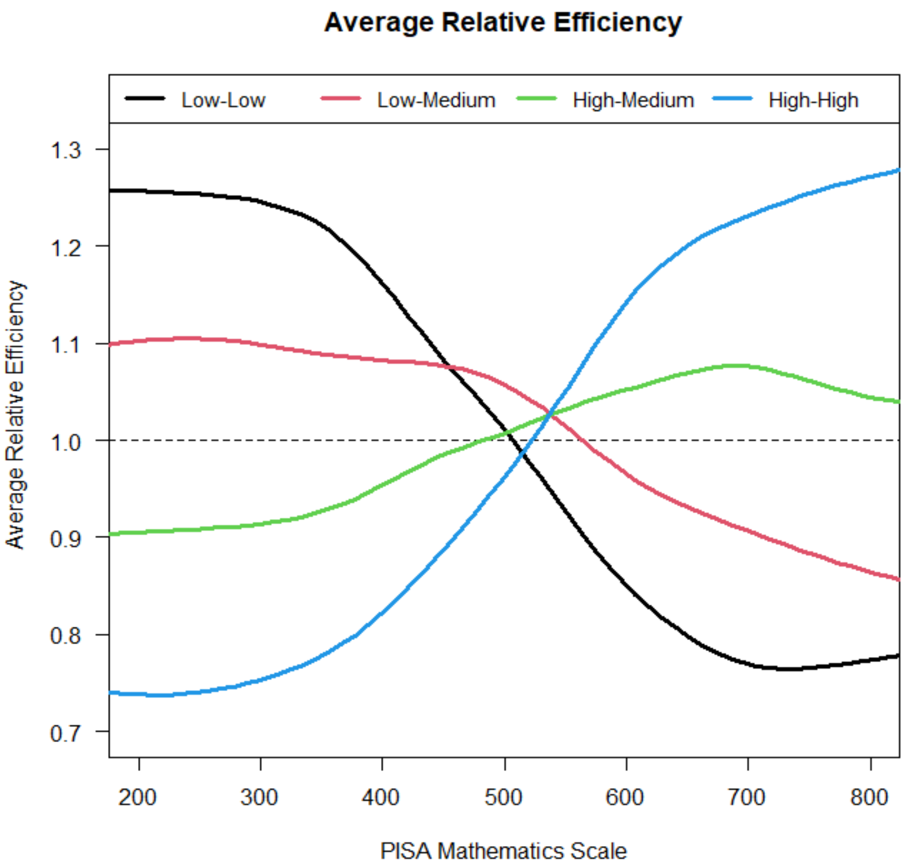
$$\begin{gathered} n_o^{\min} \leq \sum_{u=1}^{U} n_u z_{utt'} \leq n_o^{\max}, \quad \text{for all t < t',} \\ 2z_{utt'} \leq x_{ut} + x_{ut'}, \quad \text{for all u,} \\ z_{utt'} \geq x_{ut} + x_{ut'} - 1, \quad \text{for all u,} \end{gathered}$$

where $n_o^{\min}$ and $n_o^{\max}$ are the minimum and maximum number of common items, n_u is the number of items of unit *u*, $z_{utt'}$ are additional decision variables that indicate whether unit *u* is in both testlet *t* and *t'* with *t* < *t'*. The last two constraints are needed to keep the values of the decision variables consistent [see van der Linden (2005, p. 145[3])].

Across all steps of the assembly, the total number of decision variables was about 92,000 and the total number of constraints was about 174,000, too many to list all of them here. Furthermore, the assembly was an iterative process in the sense that desired constraints could not always be implemented due to availability (e.g., not enough items of a specific type) or infeasibility (i.e., no solution could be found). In the latter case, a process called feasibility relaxation was used in which weights were assigned to give higher priority to more problematic constraint violations (e.g., items being overused) and lower priority to less problematic constraint violations (e.g., content constraints) [e.g., Lundell and Kronqvist (2022[6])].

To evaluate the expected efficiency of the MSAT design, Figure 2.12 shows the average relative efficiency based on the average TIF of the MSAT paths over the average TIF of linear forms using estimated item parameters from the field trial (only international item parameters were used). Values larger than one indicate that the MSAT paths are more efficient than the linear forms. It can be seen that the MSAT paths provide more information than the linear forms when the proficiency level would match the difficulty level of the path (e.g., the curve for the low-low path exceeds one for lower proficiency values).

Figure 2.12. Average relative efficiency of MSAT paths over linear forms for PISA 2022 mathematics test design



References

Ali, U. and P. van Rijn (2016), “An evaluation of different statistical targets for assembling parallel forms in item response theory”, *Applied Psychological Measurement*, Vol. 40/3, pp. 163-179, https://doi.org/10.1177/0146621615613308. [2]

Diao, Q. and W. van der Linden (2011), “Automated test assembly using lp_solve version 5.5 in R”, *Applied Psychological Measurement*, Vol. 35/5, pp. 398-409. [4]

Lundell, A. and J. Kronqvist (2022), “Polyhedral approximation strategies for nonconvex mixed-integer nonlinear programming in SHOT”, *Journal of Global Optimization*, Vol. 82/4, pp. 863-896. [6]

van der Linden, W. (2005), *Linear Models for Optimal Test Design*, Springer, New York. [3]

van Rijn, P. et al. (2022), *Stepwise assembly for multistage adaptive testing: An application to PISA mathematics*, Presentation at IACAT conference, Frankfurt, Germany. [5]

von Davier, M. et al. (2019), “Evaluating item response theory linking and model fit for data from PISA 2000-2012”, *Assessment in Education: Principles, Policy & Practice*, Vol. 26/4, pp. 466-488. [1]

Notes

1. The mode of assessment for most of the participants was computer-based (77 CBA participants), with 4 participants implementing the PISA 2022 cycle as a paper-based survey.

2. See https://www.oecd.org/pisa/pisa-for-development/.

Chapter 2 tables

Tables	Title
Table 2.1	Number of items by domain and across cycles in the main survey
Table 2.2	Main survey computer-based MSAT paths for mathematics
Table 2.3	Main survey computer-based linear forms for mathematics

Table 2.1. Number of items by domain and across cycles in the main survey

	2000	2003	2006	2009	2012	2015	2018	2022
Reading	**129**	28	28	**131**	44	103	**245**	197
Mathematics	43	**84**	48	35	**109**	83	83	**234**
Science	45	34	**103**	53	53	**184**	115	115
Total Across Domains	217	146	179	219	206	370	443	546

Note: Bold numbers indicate the major domain in each cycle. For the 2015 and 2018 cycles, the computer-based mathematics instrument contained 82 items, while the equivalent paper-based instrument contained 83 items. This is because there was one item that was not able to be transitioned to a computer-based delivery in 2015 (the item requires students to draw on a map). The number of mathematics items in the 2022 cycle includes 74 "trend" items (i.e., items developed prior to this cycle) and 160 "new" items (i.e., items developed this cycle).

Table 2.2. Main survey computer-based MSAT paths for mathematics

MSAT_Path	Difficulty_Level	Core_Testlet	Adaptive_1_Testlet	Adaptive_2_Testlet
1	Low_Low	MTA001	MTB108	MTC203
2	Low_Low	MTA002	MTB103	MTC204
3	Low_Low	MTA003	MTB105	MTC204
4	Low_Low	MTA004	MTB105	MTC201
5	Low_Low	MTA005	MTB104	MTC202
6	Low_Low	MTA006	MTB107	MTC204
7	Low_Low	MTA007	MTB104	MTC202
8	Low_Low	MTA008	MTB108	MTC203
9	Low_Low	MTA009	MTB101	MTC201
10	Low_Low	MTA010	MTB106	MTC202
11	Low_Low	MTA011	MTB101	MTC201
12	Low_Low	MTA012	MTB103	MTC203
13	Low_Low	MTA013	MTB107	MTC202
14	Low_Low	MTA014	MTB102	MTC201
15	Low_Low	MTA015	MTB102	MTC203
16	Low_Low	MTA016	MTB106	MTC204
17	Low_Medium	MTA001	MTB108	MTC206
18	Low_Medium	MTA002	MTB103	MTC212
19	Low_Medium	MTA003	MTB105	MTC205
20	Low_Medium	MTA004	MTB105	MTC208
21	Low_Medium	MTA005	MTB104	MTC211
22	Low_Medium	MTA006	MTB107	MTC206
23	Low_Medium	MTA007	MTB104	MTC207
24	Low_Medium	MTA008	MTB108	MTC210
25	Low_Medium	MTA009	MTB101	MTC208
26	Low_Medium	MTA010	MTB106	MTC210
27	Low_Medium	MTA011	MTB101	MTC212
28	Low_Medium	MTA012	MTB103	MTC209

MSAT_Path	Difficulty_Level	Core_Testlet	Adaptive_1_Testlet	Adaptive_2_Testlet
29	Low_Medium	MTA013	MTB107	MTC211
30	Low_Medium	MTA014	MTB102	MTC209
31	Low_Medium	MTA015	MTB102	MTC205
32	Low_Medium	MTA016	MTB106	MTC207
33	High_Medium	MTA001	MTB113	MTC206
34	High_Medium	MTA002	MTB115	MTC212
35	High_Medium	MTA003	MTB110	MTC205
36	High_Medium	MTA004	MTB112	MTC211
37	High_Medium	MTA005	MTB116	MTC207
38	High_Medium	MTA006	MTB111	MTC209
39	High_Medium	MTA007	MTB114	MTC211
40	High_Medium	MTA008	MTB114	MTC210
41	High_Medium	MTA009	MTB113	MTC208
42	High_Medium	MTA010	MTB110	MTC210
43	High_Medium	MTA011	MTB115	MTC212
44	High_Medium	MTA012	MTB109	MTC206
45	High_Medium	MTA013	MTB116	MTC207
46	High_Medium	MTA014	MTB109	MTC205
47	High_Medium	MTA015	MTB111	MTC209
48	High_Medium	MTA016	MTB112	MTC208
49	High_High	MTA001	MTB113	MTC213
50	High_High	MTA002	MTB115	MTC214
51	High_High	MTA003	MTB110	MTC216
52	High_High	MTA004	MTB112	MTC213
53	High_High	MTA005	MTB116	MTC215
54	High_High	MTA006	MTB111	MTC214
55	High_High	MTA007	MTB114	MTC215
56	High_High	MTA008	MTB114	MTC214
57	High_High	MTA009	MTB113	MTC213
58	High_High	MTA010	MTB110	MTC216
59	High_High	MTA011	MTB115	MTC214
60	High_High	MTA012	MTB109	MTC215
61	High_High	MTA013	MTB116	MTC215
62	High_High	MTA014	MTB109	MTC216
63	High_High	MTA015	MTB111	MTC213
64	High_High	MTA016	MTB112	MTC216
65	Low_Low	MTB001	MTC103	MTA204
66	Low_Low	MTB002	MTC107	MTA201
67	Low_Low	MTB003	MTC101	MTA204
68	Low_Low	MTB004	MTC106	MTA203
69	Low_Low	MTB005	MTC104	MTA201
70	Low_Low	MTB006	MTC103	MTA204
71	Low_Low	MTB007	MTC105	MTA203
72	Low_Low	MTB008	MTC102	MTA203
73	Low_Low	MTB009	MTC108	MTA202
74	Low_Low	MTB010	MTC106	MTA201
75	Low_Low	MTB011	MTC108	MTA202
76	Low_Low	MTB012	MTC107	MTA201
77	Low_Low	MTB013	MTC105	MTA203
78	Low_Low	MTB014	MTC104	MTA202
79	Low_Low	MTB015	MTC101	MTA204
80	Low_Low	MTB016	MTC102	MTA202
81	Low_Medium	MTB001	MTC103	MTA212
82	Low_Medium	MTB002	MTC107	MTA206

MSAT_Path	Difficulty_Level	Core_Testlet	Adaptive_1_Testlet	Adaptive_2_Testlet
83	Low_Medium	MTB003	MTC101	MTA211
84	Low_Medium	MTB004	MTC106	MTA208
85	Low_Medium	MTB005	MTC104	MTA205
86	Low_Medium	MTB006	MTC103	MTA211
87	Low_Medium	MTB007	MTC105	MTA208
88	Low_Medium	MTB008	MTC102	MTA209
89	Low_Medium	MTB009	MTC108	MTA206
90	Low_Medium	MTB010	MTC106	MTA207
91	Low_Medium	MTB011	MTC108	MTA209
92	Low_Medium	MTB012	MTC107	MTA207
93	Low_Medium	MTB013	MTC105	MTA205
94	Low_Medium	MTB014	MTC104	MTA210
95	Low_Medium	MTB015	MTC101	MTA212
96	Low_Medium	MTB016	MTC102	MTA210
97	High_Medium	MTB001	MTC113	MTA211
98	High_Medium	MTB002	MTC114	MTA208
99	High_Medium	MTB003	MTC112	MTA212
100	High_Medium	MTB004	MTC113	MTA212
101	High_Medium	MTB005	MTC110	MTA205
102	High_Medium	MTB006	MTC114	MTA208
103	High_Medium	MTB007	MTC109	MTA211
104	High_Medium	MTB008	MTC115	MTA207
105	High_Medium	MTB009	MTC110	MTA209
106	High_Medium	MTB010	MTC111	MTA209
107	High_Medium	MTB011	MTC115	MTA206
108	High_Medium	MTB012	MTC116	MTA210
109	High_Medium	MTB013	MTC109	MTA206
110	High_Medium	MTB014	MTC116	MTA210
111	High_Medium	MTB015	MTC112	MTA205
112	High_Medium	MTB016	MTC111	MTA207
113	High_High	MTB001	MTC113	MTA213
114	High_High	MTB002	MTC114	MTA215
115	High_High	MTB003	MTC112	MTA216
116	High_High	MTB004	MTC113	MTA214
117	High_High	MTB005	MTC110	MTA214
118	High_High	MTB006	MTC114	MTA215
119	High_High	MTB007	MTC109	MTA216
120	High_High	MTB008	MTC115	MTA214
121	High_High	MTB009	MTC110	MTA213
122	High_High	MTB010	MTC111	MTA215
123	High_High	MTB011	MTC115	MTA216
124	High_High	MTB012	MTC116	MTA213
125	High_High	MTB013	MTC109	MTA216
126	High_High	MTB014	MTC116	MTA215
127	High_High	MTB015	MTC112	MTA214
128	High_High	MTB016	MTC111	MTA213
129	Low_Low	MTC001	MTA103	MTB201
130	Low_Low	MTC002	MTA101	MTB202
131	Low_Low	MTC003	MTA107	MTB202
132	Low_Low	MTC004	MTA105	MTB201
133	Low_Low	MTC005	MTA101	MTB204
134	Low_Low	MTC006	MTA104	MTB201
135	Low_Low	MTC007	MTA108	MTB202
136	Low_Low	MTC008	MTA104	MTB203

MSAT_Path	Difficulty_Level	Core_Testlet	Adaptive_1_Testlet	Adaptive_2_Testlet
137	Low_Low	MTC009	MTA103	MTB203
138	Low_Low	MTC010	MTA102	MTB203
139	Low_Low	MTC011	MTA106	MTB204
140	Low_Low	MTC012	MTA107	MTB202
141	Low_Low	MTC013	MTA106	MTB204
142	Low_Low	MTC014	MTA105	MTB201
143	Low_Low	MTC015	MTA108	MTB204
144	Low_Low	MTC016	MTA102	MTB203
145	Low_Medium	MTC001	MTA103	MTB210
146	Low_Medium	MTC002	MTA101	MTB212
147	Low_Medium	MTC003	MTA107	MTB207
148	Low_Medium	MTC004	MTA105	MTB211
149	Low_Medium	MTC005	MTA101	MTB208
150	Low_Medium	MTC006	MTA104	MTB209
151	Low_Medium	MTC007	MTA108	MTB211
152	Low_Medium	MTC008	MTA104	MTB208
153	Low_Medium	MTC009	MTA103	MTB210
154	Low_Medium	MTC010	MTA102	MTB205
155	Low_Medium	MTC011	MTA106	MTB206
156	Low_Medium	MTC012	MTA107	MTB209
157	Low_Medium	MTC013	MTA106	MTB212
158	Low_Medium	MTC014	MTA105	MTB205
159	Low_Medium	MTC015	MTA108	MTB207
160	Low_Medium	MTC016	MTA102	MTB206
161	High_Medium	MTC001	MTA112	MTB205
162	High_Medium	MTC002	MTA115	MTB210
163	High_Medium	MTC003	MTA110	MTB209
164	High_Medium	MTC004	MTA113	MTB207
165	High_Medium	MTC005	MTA109	MTB206
166	High_Medium	MTC006	MTA109	MTB212
167	High_Medium	MTC007	MTA112	MTB208
168	High_Medium	MTC008	MTA114	MTB207
169	High_Medium	MTC009	MTA115	MTB208
170	High_Medium	MTC010	MTA110	MTB205
171	High_Medium	MTC011	MTA113	MTB211
172	High_Medium	MTC012	MTA116	MTB209
173	High_Medium	MTC013	MTA114	MTB206
174	High_Medium	MTC014	MTA116	MTB210
175	High_Medium	MTC015	MTA111	MTB212
176	High_Medium	MTC016	MTA111	MTB211
177	High_High	MTC001	MTA112	MTB215
178	High_High	MTC002	MTA115	MTB214
179	High_High	MTC003	MTA110	MTB216
180	High_High	MTC004	MTA113	MTB214
181	High_High	MTC005	MTA109	MTB216
182	High_High	MTC006	MTA109	MTB215
183	High_High	MTC007	MTA112	MTB213
184	High_High	MTC008	MTA114	MTB214
185	High_High	MTC009	MTA115	MTB214
186	High_High	MTC010	MTA110	MTB216
187	High_High	MTC011	MTA113	MTB213
188	High_High	MTC012	MTA116	MTB216
189	High_High	MTC013	MTA114	MTB213
190	High_High	MTC014	MTA116	MTB215

MSAT_Path	Difficulty_Level	Core_Testlet	Adaptive_1_Testlet	Adaptive_2_Testlet
191	High_High	MTC015	MTA111	MTB213
192	High_High	MTC016	MTA111	MTB215

Where:
MT = Math Testlet
A-B-C = Set
0-1-2 (the single digit immediately to the right of the set letter) = Stage
0 = core, 1 = adaptive stage 1, 2 = adaptive stage 2
01-16 (the last two digits on the right) = testlet number

Examples:

MTA008	**MT**	**A**	**0**	**08**
	Math Testlet	set A	core stage	testlet 08
MTB214	**MT**	**B**	**2**	**14**
	Math Testlet	set B	adaptive stage 2	testlet 14

Table 2.3. Main survey computer-based linear forms for mathematics

Linear_Form	Core_Testlet_1	Core_Testlet_2	Core_Testlet_3
1	MTA001	MTB010	MTC015
2	MTA002	MTB014	MTC010
3	MTA003	MTB001	MTC013
4	MTA004	MTB011	MTC014
5	MTA005	MTB005	MTC005
6	MTA006	MTB008	MTC004
7	MTA007	MTB016	MTC012
8	MTA008	MTB007	MTC016
9	MTA009	MTB006	MTC009
10	MTA010	MTB009	MTC011
11	MTA011	MTB004	MTC008
12	MTA012	MTB013	MTC006
13	MTA013	MTB002	MTC002
14	MTA014	MTB003	MTC001
15	MTA015	MTB012	MTC007
16	MTA016	MTB015	MTC003
17	MTB001	MTC001	MTA010
18	MTB002	MTC005	MTA002
19	MTB003	MTC008	MTA004
20	MTB004	MTC003	MTA015
21	MTB005	MTC014	MTA007
22	MTB006	MTC010	MTA011
23	MTB007	MTC016	MTA005
24	MTB008	MTC006	MTA012
25	MTB009	MTC011	MTA014
26	MTB010	MTC015	MTA001
27	MTB011	MTC013	MTA016
28	MTB012	MTC002	MTA009
29	MTB013	MTC009	MTA006
30	MTB014	MTC007	MTA003
31	MTB015	MTC004	MTA008
32	MTB016	MTC012	MTA013
33	MTC001	MTA012	MTB002
34	MTC002	MTA005	MTB007

Linear_Form	Core_Testlet_1	Core_Testlet_2	Core_Testlet_3
35	MTC003	MTA015	MTB013
36	MTC004	MTA008	MTB015
37	MTC005	MTA002	MTB004
38	MTC006	MTA006	MTB008
39	MTC007	MTA016	MTB012
40	MTC008	MTA004	MTB011
41	MTC009	MTA010	MTB006
42	MTC010	MTA009	MTB009
43	MTC011	MTA014	MTB014
44	MTC012	MTA013	MTB016
45	MTC013	MTA007	MTB003
46	MTC014	MTA003	MTB005
47	MTC015	MTA001	MTB010
48	MTC016	MTA011	MTB001

Where:
MT = Math Testlet
A-B-C = Set
0 (the single digit immediately to the right of the set letter) = Stage
Only core testlets were used with the linear design
01-16 (the last two digits on the right) = testlet number

Annex 2.A. Core testlet cut scores

Annex Table 2.A.1. Core testlet cut scores

Core Testlet	Core Cut Score Low-High	Max. Machine Score	Max. Total Score
MTA001	6	10	13
MTA002	5	11	13
MTA003	6	11	13
MTA004	6	9	12
MTA005	6	11	13
MTA006	6	9	13
MTA007	6	10	12
MTA008	6	9	12
MTA009	5	11	13
MTA010	5	11	13
MTA011	5	10	13
MTA012	6	11	13
MTA013	5	9	12
MTA014	6	11	13
MTA015	6	11	13
MTA016	6	9	12
MTB001	6	9	12
MTB002	6	9	13
MTB003	6	10	12
MTB004	6	11	13
MTB005	6	11	13
MTB006	6	9	12
MTB007	6	9	12
MTB008	5	11	13
MTB009	6	10	13
MTB010	6	11	13
MTB011	6	11	13
MTB012	6	11	13
MTB013	5	11	13
MTB014	6	11	13
MTB015	6	9	12
MTB016	6	10	12
MTC001	6	10	12
MTC002	6	11	13
MTC003	6	11	13
MTC004	5	11	13

Core Testlet	Core Cut Score Low-High	Max. Machine Score	Max. Total Score
MTC005	6	10	12
MTC006	6	11	13
MTC007	6	9	11
MTC008	6	12	14
MTC009	5	11	13
MTC010	6	10	12
MTC011	5	10	12
MTC012	5	9	13
MTC013	5	10	12
MTC014	6	10	13
MTC015	5	10	12
MTC016	6	11	13

Annex 2.B. Adaptive testlet cut scores

Annex Table 2.B.1. Adaptive testlet cut scores

Core Testlet	Adaptive 1 Testlet	Adaptive 1 Cut Score Low-Medium	Adaptive 1 Cut Score Medium-High	Max. Machine Score	Max. Total Score
MTA001	MTB108	9	99	20	26
MTA001	MTB113	99	12	19	25
MTA002	MTB103	9	99	21	25
MTA002	MTB115	99	12	20	25
MTA003	MTB105	10	99	21	25
MTA003	MTB110	99	12	21	26
MTA004	MTB105	10	99	19	24
MTA004	MTB112	99	11	19	24
MTA005	MTB104	11	99	22	26
MTA005	MTB116	99	12	20	25
MTA006	MTB107	10	99	20	26
MTA006	MTB111	99	13	20	26
MTA007	MTB104	11	99	21	25
MTA007	MTB114	99	12	21	25
MTA008	MTB108	9	99	19	25
MTA008	MTB114	99	12	20	25
MTA009	MTB101	10	99	22	26
MTA009	MTB113	99	11	20	25
MTA010	MTB106	10	99	21	26
MTA010	MTB110	99	11	21	26
MTA011	MTB101	10	99	21	26
MTA011	MTB115	99	12	19	25
MTA012	MTB103	10	99	21	25
MTA012	MTB109	99	14	22	26
MTA013	MTB107	10	99	20	25
MTA013	MTB116	99	11	18	24
MTA014	MTB102	10	99	21	25
MTA014	MTB109	99	13	22	26
MTA015	MTB102	11	99	21	25
MTA015	MTB111	99	13	22	26
MTA016	MTB106	11	99	19	25
MTA016	MTB112	99	11	19	24
MTB001	MTC103	10	99	19	24
MTB001	MTC113	99	12	20	25
MTB002	MTC107	11	99	20	26
MTB002	MTC114	99	12	19	25
MTB003	MTC101	10	99	20	24

Core Testlet	Adaptive 1 Testlet	Adaptive 1 Cut Score Low-Medium	Adaptive 1 Cut Score Medium-High	Max. Machine Score	Max. Total Score
MTB003	MTC112	99	12	20	24
MTB004	MTC106	9	99	21	26
MTB004	MTC113	99	12	22	26
MTB005	MTC104	11	99	22	26
MTB005	MTC110	99	13	21	25
MTB006	MTC103	10	99	19	24
MTB006	MTC114	99	12	19	24
MTB007	MTC105	10	99	19	25
MTB007	MTC109	99	12	19	25
MTB008	MTC102	9	99	22	26
MTB008	MTC115	99	11	21	25
MTB009	MTC108	10	99	20	25
MTB009	MTC110	99	13	20	25
MTB010	MTC106	9	99	21	26
MTB010	MTC111	99	11	21	26
MTB011	MTC108	10	99	21	25
MTB011	MTC115	99	13	21	25
MTB012	MTC107	11	99	22	26
MTB012	MTC116	99	12	21	26
MTB013	MTC105	10	99	21	26
MTB013	MTC109	99	12	21	26
MTB014	MTC104	11	99	22	26
MTB014	MTC116	99	13	21	26
MTB015	MTC101	10	99	19	24
MTB015	MTC112	99	12	19	24
MTB016	MTC102	10	99	21	25
MTB016	MTC111	99	11	20	25
MTC001	MTA103	10	99	20	24
MTC001	MTA112	99	13	21	25
MTC002	MTA101	10	99	21	25
MTC002	MTA115	99	13	21	25
MTC003	MTA107	9	99	21	27
MTC003	MTA110	99	12	21	26
MTC004	MTA105	9	99	22	26
MTC004	MTA113	99	12	22	26
MTC005	MTA101	9	99	20	24
MTC005	MTA109	99	12	21	25
MTC006	MTA104	11	99	21	25
MTC006	MTA109	99	13	22	26
MTC007	MTA108	10	99	19	23
MTC007	MTA112	99	13	20	24
MTC008	MTA104	10	99	22	26
MTC008	MTA114	99	12	22	27
MTC009	MTA103	9	99	21	25

Core Testlet	Adaptive 1 Testlet	Adaptive 1 Cut Score Low-Medium	Adaptive 1 Cut Score Medium-High	Max. Machine Score	Max. Total Score
MTC009	MTA115	99	12	21	25
MTC010	MTA102	9	99	19	25
MTC010	MTA110	99	12	20	25
MTC011	MTA106	10	99	21	25
MTC011	MTA113	99	12	21	25
MTC012	MTA107	9	99	19	27
MTC012	MTA116	99	11	21	25
MTC013	MTA106	10	99	21	25
MTC013	MTA114	99	12	20	25
MTC014	MTA105	11	99	21	26
MTC014	MTA116	99	13	22	25
MTC015	MTA108	9	99	20	24
MTC015	MTA111	99	12	21	25
MTC016	MTA102	9	99	20	26
MTC016	MTA111	99	13	22	26

Note: 99 = not applicable.