Technical Cooperation Group
9th meeting
22-24 November 2022

WG/GAML/16

PISA Household Survey Module (PISA-HSM)

Organisation for Economic Co-operation and Development (OECD)[1]

Global Alliance to Monitor Learning (GAML)

23 November 2022

[1] This document was prepared by Tomoya Okubo, Francesco Avvisati and Michael Ward (OECD, 11-23-2022)

UIS and TCG

7th meeting of the Global Alliance to Monitor Learning

PISA Household Survey Module (HSM)

Tomoya Okubo, Francesco Avvisati and Michael Ward
OECD
11-23-2022

PISA Household Survey Module (PISA-HSM)

Table of Contents

Table of Figures

Table of Tables

1. Introduction

1.1. Background

The first PISA-based test administered in the context of a household survey was the PISA-D out-of-school assessment ("Strand C"). Strand C used items from the domains of reading and mathematics that were originally included in the regular (in-school) PISA for Development assessment, but reduced the proportion of items measuring higher levels of proficiency to better match the test information to the expected proficiency of out-of-school children.

Building on the Strand C instruments and experience, the OECD has developed a PISA household survey module (PISA-HSM) designed to assess solely whether 14-to-16-year-olds have acquired "minimum levels of proficiency in reading and mathematics" as defined by PISA and agreed with the UN system for SDG 4 monitoring. PISA-HSM has been developed through a process that has tailored the assessment to a) the measurement goal and b) the operational constraints of a household administration. PISA-HSM may be administered as a paper based assessment or on a tablet computer (as per Strand C).

This document describes how a fit-for-purpose module has been built based on the items originally used in the PISA for Development assessment. It uses the regular PISA-D test and the Strand C tests as a benchmark against which the validity and reliability of the resulting test is evaluated. Two versions of PISA-HSM have been prepared – a 30 minute test and a 45 minute test.

2. PISA for Development

2.1. Item clusters of PISA-D

In PISA for Development (PISA-D), there are three domains in the cognitive tests; mathematics (MATH), reading (READ), and science (SCIE). In each domain, items are assigned to one of four item-clusters. Each item cluster is designed as a 30 minutes block. The cluster information of MATH and READ are shown in Table 1 and Table 2, respectively. Please note that SCIE is not discussed in this report.

In MATH, each cluster composes 16 items (10-13 units) taken from the other OECD surveys such as PISA, PISA for Schools, and PIAAC. READ composes 16-17 items (5-6 units) selected from PISA, PISA for Schools, LAMP (by UNESCO), and PIAAC.

Students take two item-clusters in a domain, and they take two domains in a test. Therefore, a student takes a 120 minutes test containing four 30 minutes item-clusters. Response formats of the PISA-D items are keyword (KW), multiple-choice (MCQ), True/False type (X-type), and Open-ended type (Open-end).

Table 1. Item clusters of MATH in PISA-D

Domain	Unit	Item	Category 1	Category 2	Source
M1	13	16	Space and Shape, Change and Relationships, Uncertainty and data	Interpret, Formulate, Employ	PISA 2015 Trend, PISA for Schools, PIAAC
M2	11	16	Quantity, Uncertainty and Data, Space and Shape	Employ, Interpret, Formulate	PISA 2015 Trend, PIAAC, PISA for Schools
M3	10	16	Space and Shape, Quantity, Uncertainty and data	Uncertainty and Data, Employ, Formulate	PISA 2015 Trend, PISA for Schools, PIAAC
M4	10	16	Quantity, Space and Shape, Uncertainty and Data	Interpret, Formulate, Employ	PIAAC, PISA 2015 Trend, PISA for Schools

Table 2. Item clusters of READ in PISA-D

Domain	Unit	Item	Category 1	Category 2	Source
R1	5	17	Access and retrieve, Integrate and interpret, Reflect and evaluate	Personal, Educational, Public	PIAAC , PISA 2015 Trend,
R2	5	16	Integrate and interpret, Access and retrieve, Reflect and evaluate	Personal, Educational, Public	LAMP, PISA for Schools, PISA 2015 Trend
R3	6	16	Access and retrieve, Reflect and evaluate, Integrate and interpret	Occupational, Personal, Educational	PIAAC, PISA 2015 Trend, PISA for Schools
R4	6	17	Integrate and interpret, Reflect and evaluate, Access and retrieve	Public, Educational, Personal	LAMP, PISA 2015 Trend, PIAAC

2.2. Booklet design of PISA-D

In PISA-D, each booklet consists of two domains, and each domain contains two item-clusters. Table 3 shows the booklet design of PISA-D. It also shows the total number of items(units) in each domain (except SCIE). The number of items in each domain (30 minutes + 30 minutes) is 32-34, and in total, students take about 65 items in 120 minutes.

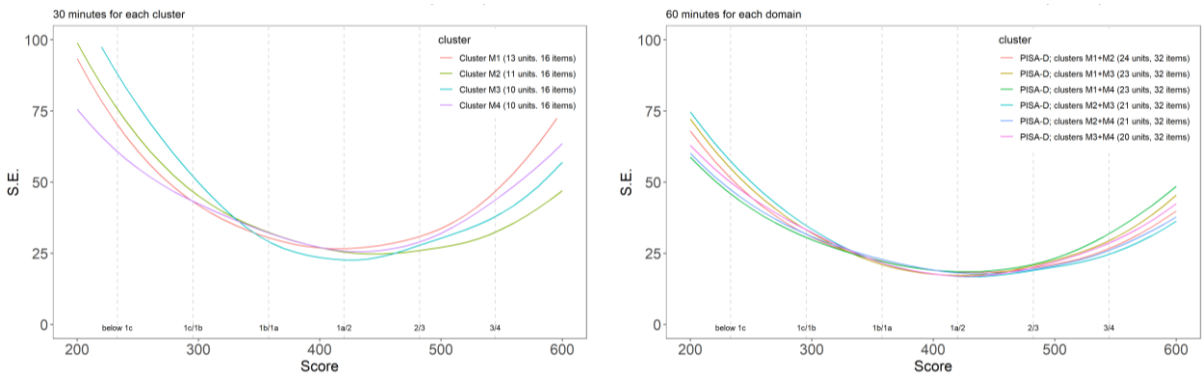
Table 3. Booklet design of PISA-D

Booklet #	1st session	2nd session	3rd session	4th session	MATH	READ
1	R1	R2	S1	S2	0	10(33)
2	S2	S3	R2	R3	0	11(32)
3	R3	R4	S3	S4	0	12(33)
4	S4	S1	R4	R1	0	11(34)
5	S1	S2	M1	M2	24(32)	0
6	M2	M3	S2	S3	22(32)	0
7	S3	S4	M3	M4	20(32)	0
8	M4	M1	S4	S1	23(32)	0
9	M1	M2	R1	R2	24(32)	10(33)
10	R2	R3	M2	M3	21(32)	11(32)
11	M3	M4	R3	R4	20(32)	12(33)
12	R4	R1	M4	M1	23(32)	11(34)

2.3. Reliability of PISA-D

The standard error functions of the four MATH item-clusters (left-side) and the sets of two MATH item-clusters (right-side) are shown in Figure 1. It shows that item-clusters in MATH domain are almost equivalent to each other in terms of reliability of the measures. Furthermore, it indicates that the sets of two item-clusters (i.e., booklets) are also equivalent. The MATH tests of PISA-D cover a wide range of proficiency levels.

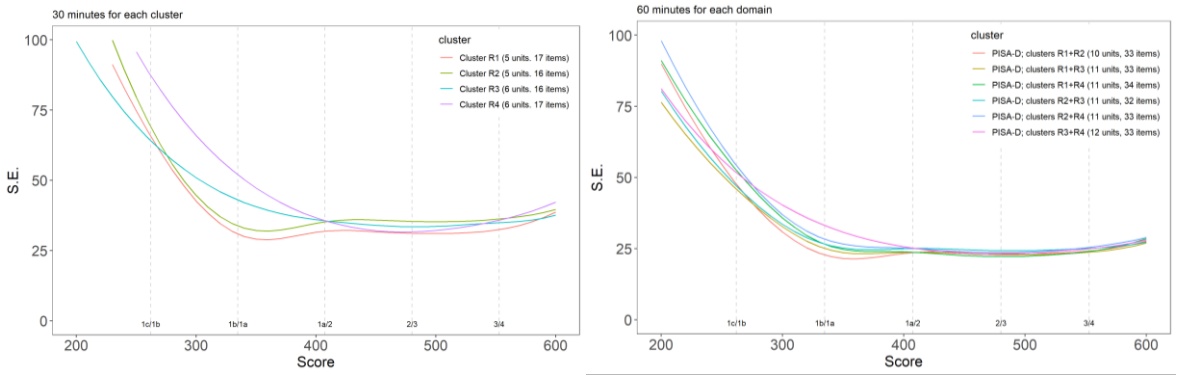
Figure 1 Standard Error of MATH item clusters (left-side) and MATH booklets (right-side) in PISA-D



The standard error of the four READ item-clusters (left-side) and the sets of two item-clusters (right-side) are shown in Figure 2. **Error! Reference source not found.** shows that the item-clusters are almost equivalent in terms of reliability of scores, especially for scores between 400 and 600. Clusters R1 and R2 are slightly different from the others in terms of reliability, especially for scores below 400. The figures indicate that the

READ booklets of PISA-D cover a wide range of proficiency levels, especially students whose proficiency levels are 1a and above.

Figure 2 Standard Error of READ item clusters (left-side) and READ booklets (right-side) in PISA-D



3. PISA Strand-C

3.1. Item clusters of PISA Strand-C

PISA for Development Strand C (PISA Strand-C) is a short version of PISA-D in which students take MATH and READ cognitive tests[1] in 45 minutes. The items of PISA Strand-C are selected from PISA-D. Table 4 and Table 5 show item cluster information on MATH and READ, respectively. In PISA Strand-C, the response formats of the items are mainly keyword, multiple-choice, and X-type.

Table 4. Item clusters of MATH in PISA Strand-C

Domain	Unit	Item	Category 1	Category 2	Source
Mcore	5	5	Space and Shape, Quantity, Uncertainty and data	Interpret, Employ	PISA 2015 Trend, PISA for Schools, PIAAC
M1	9	11	Quantity, Uncertainty and Data, Space and Shape, Change and Relationships	Employ, Interpret, Formulate	PISA 2015 Trend, PIAAC, PISA for Schools
M2	7	9	Change and Relationships, Quantity, Uncertainty and data	Interpret, Employ, Formulate	PISA 2015 Trend, PISA for Schools, PIAAC
M3	8	10	Quantity, Space and Shape, Change and Relationships	Interpret, Formulate, Employ	PIAAC, PISA 2015 Trend, PISA for Schools

[1] we do not consider the reading component (RC) here.

Table 5. Item clusters of READ in PISA Strand-C

Domain	Unit	Item	Category 1	Category 2	Source
Rcore	3	5	Access and retrieve, Integrate and interpret	Personal, Occupational	PIAAC , PISA 2015 Trend, LAMP
R1	3	6	Integrate and interpret, Access and retrieve, Reflect and evaluate	Personal, Educational	PISA for Schools, PISA 2015 Trend
R2	3	5	Access and retrieve, Integrate and interpret	Public, Educational	PISA for Schools
R3	2	6	Integrate and interpret	Public, Educational	LAMP, PISA 2015 Trend

3.2. Booklet design of PISA Strand-C

In PISA Strand-C, each booklet contains both MATH items and READ items. Each domain consists of the core cluster (i.e., Rcore and Mcore) and one item-cluster or two item-clusters. Table 6 shows the booklet design of PISA Strand-C. There are 12 booklets in PISA Strand-C, and students take either of 1) 14-16 MATH items (12-14 units) and 16-17 READ items (8-9 units), or 2) 24-26 MATH items (20-22 units) and 10-11 READ items (5-6 units). Students take more than 30 items in a 45 minutes test session; therefore, the number of items students take in the test is more in PISA Strand-C than PISA-D.

Table 6. Booklet design of PISA-D Strand C

Booklet #	Core session	1st session	2nd session	3rd session	4th session	MATH	READ
1	Rcore+Mcore	RC-A	R1	R2	M1	14(16)	9(16)
2	Rcore+Mcore	RC-B	R2	R3	M2	12(14)	8(16)
3	Rcore+Mcore	RC-C	R3	R1	M3	13(15)	8(17)
4	Rcore+Mcore	M1	RC-B	R1	R3	14(16)	8(17)
5	Rcore+Mcore	M2	RC-C	R2	R1	12(14)	9(16)
6	Rcore+Mcore	M3	RC-A	R3	R2	13(15)	8(16)
7	Rcore+Mcore	M1	M2	RC-B	R1	21(25)	6(11)
8	Rcore+Mcore	M2	M3	RC-C	R2	20(24)	6(10)
9	Rcore+Mcore	M3	M1	RC-A	R3	22(26)	5(11)
10	Rcore+Mcore	RC-A	R1	M1	M3	22(26)	6(11)
11	Rcore+Mcore	RC-B	R2	M2	M1	21(25)	6(10)
12	Rcore+Mcore	RC-C	R3	M3	M2	20(24)	5(11)

RC represent Reading Component

3.3. Reliability of PISA Strand-C

The standard error functions of the four MATH item-clusters (left-side) and the MATH booklets[2] (right-side) are shown in Figure 3. The core cluster is more discriminative than the other MATH item-clusters in levels 1c and below. The other MATH item-clusters are overlapping each other, meaning they are equivalent in terms of reliability. The standard errors of the MATH booklets around the thresholds of levels 1a/ 2 are ranged from 20 to 35. The standard error is for a single student score; therefore, the standard error of group average will be much less than the numbers.

Figure 3 Standard Error of MATH item clusters (left-side) and MATH booklets (right-side) in PISA Strand-C

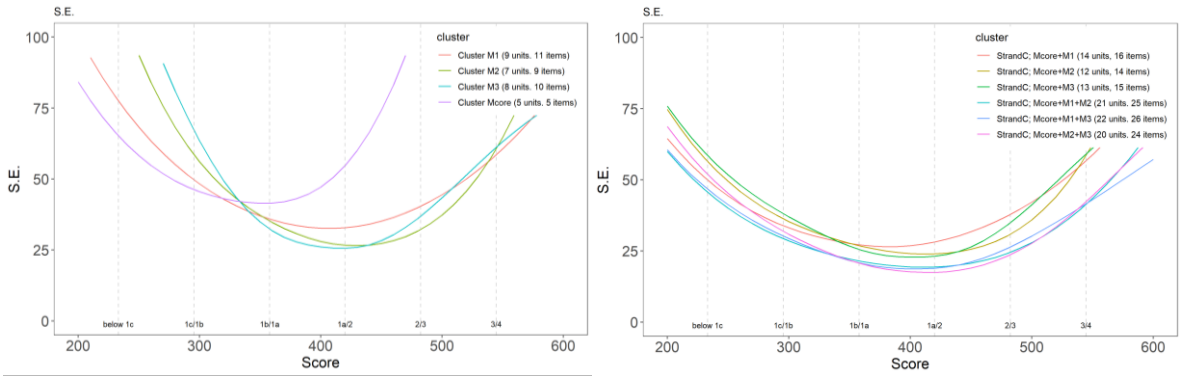
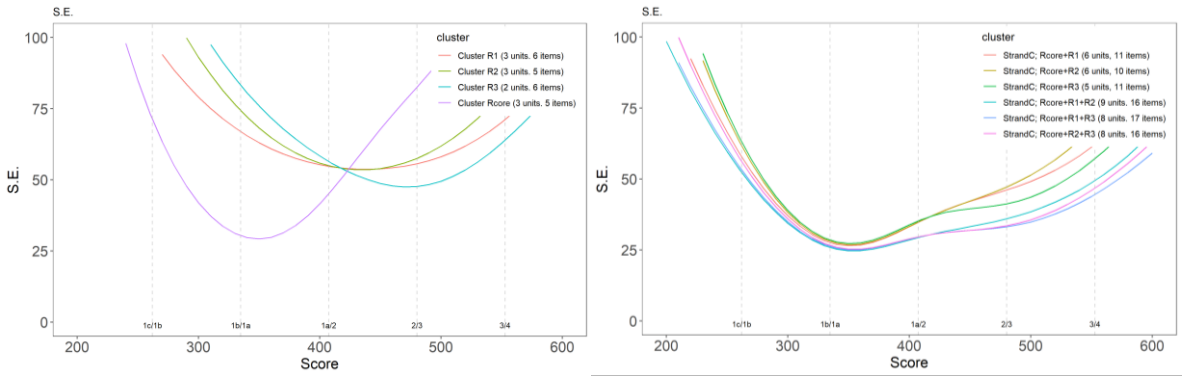


Figure 4 shows the standard error functions of the READ item-clusters (left-side) and the booklets (right-side). Like MATH clusters, the READ core cluster (i.e., Rcore) focuses on low proficiency levels comparing to the other READ clusters. The READ score's standard errors at 1a/2 levels are ranged from 30 to 40, depending on booklets. The READ booklets discriminate better in 1b/1a levels than in 1a/2 levels. The reliabilities of the scores in MATH and READ are different; the MATH scores are more reliable than the READ scores in PISA Strand-C.

Figure 4 Standard Error of READ item clusters (left-side) and READ booklets (right-side) in PISA Strand-C



[2] Here, "booklet" means sets of item-clusters. "MATH booklet" means all the MATH items in a booklet.

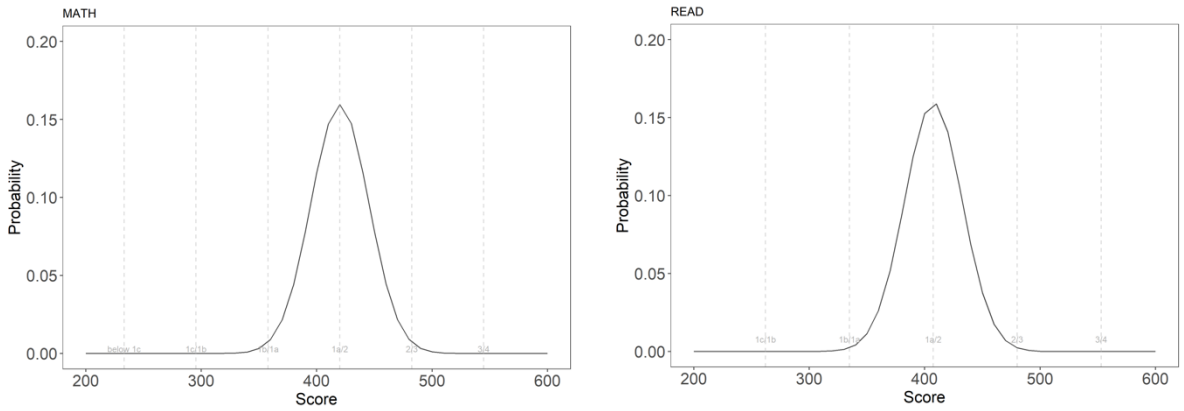
4. PISA Household Survey Module (PISA-HSM)

4.1. Purpose

The goal is to build a short test for administration as part of a household survey with the sole purpose of measuring whether 14-to-16-year-olds have reached minimum levels of proficiency in reading and mathematics as per the end of lower secondary education benchmark (SDG 4.1 target and SDG 4.1.1.c global indicator). The test shall provide valid evidence that respondents are above or below the lower limit of Level 2 proficiency on the PISA scale.

Figure 5 shows the distribution of the student proficiency levels that is assumed when assembling the MATH (left-side) tests and the READ (right-side) tests. The distribution forms normal distribution with parameters $N(420.07, 25.0)$ in MATH, and $N(407.47, 25.0)$ in READ.

Figure 5 Distributions of target samples; MATH (left-side) and READ (right-side)



4.2. PISA-HSM (30 minutes version)

Item cluster of PISA-HSM (30 mins version)

The items for PISA-HSM (30 mins) were selected in view of maximizing the reliability of measurement around the cutoff between Level 1a and Level 2, while ensuring representation of all major framework categories and with a view to including only questions that could be scored easily. The following conditions were applied when assembling items;

1) Items should be selected either from PISA trend items or PISA for Schools items.
2) Item response formats should be either keyword, multiple-choice, or X-type.
3) The booklets should cover all the subdomains.
4) Items should have enough discrimination power around the target proficiency levels shown in Figure 5. The weighted average of the information function should exceed 0.2.
5) The number of items (and units) should be
    a. Less than 7 (7 units) for MATH so that students will not take more than 10 minutes in MATH domain.
    b. Less than 14 (7 units) for READ so that students will not take more than 20 minutes in READ domain.
6) Items in a unit can be excluded from the unit if they do not satisfy the conditions mentioned above.

Table 7 and Table 8 show the selected MATH and READ items (units) for PISA-HSM(30 mins). The selected 6 MATH items (5 units) cover 68.8% of reliability against PISA-D MATH booklets and 96.6% against PISA Strand-C booklets. Here, the reliability of the MATH booklet is compared to PISA-D and PISA Strand-C MATH 1-cluster booklets, not MATH 2-clusters booklets. On the other hand, the reliability of the READ booklet is compared to READ 2-cluster booklets, not 1-cluster booklets.

For READ, 11 items (6 units) were selected from the PISA-D items. PISA-HSM (30 mins) covers 65.9% of reliability against PISA-D booklets and 92.5% against PISA Strand-C booklets. Please note that PISA-D booklets are 120 minutes long.

Table 7. Selected MATH items for PISA-HSM (30 minutes version)

Item	Unit	Category 1	Category 2	Source	Format
PM033Q01	A View With a Room	Space and Shape	Interpret	PISA 2015 Trend	MCQ
PM5104Q01	Baby Growth	Uncertainty and data	Formulate	PISA for Schools	MCQ
PM5124Q1A	Concrete Path	Change and relationships	Formulate	PISA for Schools	KW
PM5142Q01A	Machu Picchu	Quantity	Interpret	PISA for Schools	KW
PM5142Q02A	Machu Picchu	Quantity	Employ	PISA for Schools	KW
PM5169Q01*	Shoe Sizes	Change and relationships	Employ	PISA for Schools	KW

Note: Items highlighted in yellow are still in use in PISA 2022 mathematics (CBA). Items marked by an asterisk were *not* used in Strand C. All items are still in use in the PISA 2022 PBA assessment.

Table 8. Selected READ items for PISA-HSM (30 minutes version)

Item	Unit	Category 1	Category 2	Source	Format
PR101Q02	Rhino	Integrate and interpret	Public	PISA 2015 Trend	MCQ
PR101Q03	Rhino	Reflect and evaluate	Public	PISA 2015 Trend	MCQ
PR101Q04	Rhino	Integrate and interpret	Public	PISA 2015 Trend	MCQ
PR104Q01*	Telephone	Access and retrieve	Public	PISA 2015 Trend	KW
PR420Q02	Children's Futures	Access and retrieve	Educational	PISA 2015 Trend	KW
PR420Q09	Children's Futures	Access and retrieve	Educational	PISA 2015 Trend	KW
PR432Q01*	About a book	Integrate and interpret	Personal	PISA 2015 Trend	KW
PR432Q06A*	About a book	Integrate and interpret	Personal	PISA 2015 Trend	X
PR446Q03	Job Vacancy	Access and retrieve	Occupational	PISA 2015 Trend	KW
PR6015Q1A	Kenya Tourism	Reflect and evaluate	Educational	PISA for Schools	MCQ
PR6015Q6A	Kenya Tourism	Access and retrieve	Educational	PISA for Schools	MCQ

Note: Items highlighted in yellow are still in use in PISA 2022 mathematics (CBA). Items marked by an asterisk were *not* used in Strand C. All items are still in use in the PISA 2022 PBA assessment.

Booklet design of PISA-HSM (30 minutes version)

Two types of booklet designs were considered in PISA-HSM (30 mins). Table 9 shows item clusters of the tests. The READ items were split into R1 cluster and R2 cluster, while all MATH items were assigned to M1 cluster. Table 10 shows two booklet patterns; Pattern A and Pattern B. In Pattern A, two booklets were assembled, as shown in the table. The cluster M1 was placed at the first block in both booklets. In Pattern B, cluster M1 was fixed at the second block. Both patterns try to avoid the risk of not completing the MATH items in the cognitive test session.

Table 9 Item clusters of PISA-HSM (30 minutes version)

	Cluster M1		Cluster R1		Cluster R2	
1	PM033Q01	A View With a Room	PR101Q02	Rhino	PR432Q01	About a book
2	PM5104Q01	Baby Growth	PR101Q03	Rhino	PR432Q06A	About a book
3	PM5124Q1A	Concrete Path	PR101Q04	Rhino	PR446Q03	Job Vacancy
4	PM5142Q01A	Machu Picchu	PR104Q01	Telephone	PR6015Q1A	Kenya Tourism
5	PM5142Q02A	Machu Picchu	PR420Q02	Children's Futures	PR6015Q6A	Kenya Tourism
6	PM5169Q01	Shoe Sizes	PR420Q09	Children's Futures		

Table 10 Booklet design of PISA-HSM (30 minutes version)

	Pattern A		Pattern B	
	Booklet #1	**Booklet #2**	**Booklet #1**	**Booklet #2**
1st blcok	M1	M1	R1	R2
2nd block	R1	R2	M1	M1
3rd block	R2	R1	R2	R1

Reliability of PISA-HSM (30 mins)

The standard error function of MATH in PISA-HSM (30 mins) is shown in Figure 6. For the purpose of comparisons, the standard error functions of the PISA-D booklets are also shown in dashed lines. Figure 7 shows the standard error function of PISA-HSM (30 mins) together with that of PISA Strand-C booklets. The figures show that PISA-HSM (30 mins) focuses on proficiency levels around 1a and 2.

The most remarkable point here is that the number of items of PISA-HSM (30 mins) is only 20% comparing to PISA-D. Still, it keeps about 60% of the reliability. Further, it keeps 90% of the reliability of PISA Strand-C with only 40% of the number of items.

Figure 6 Standard Error of MATH scores; PISA-HSM (30 mins) vs PISA-D

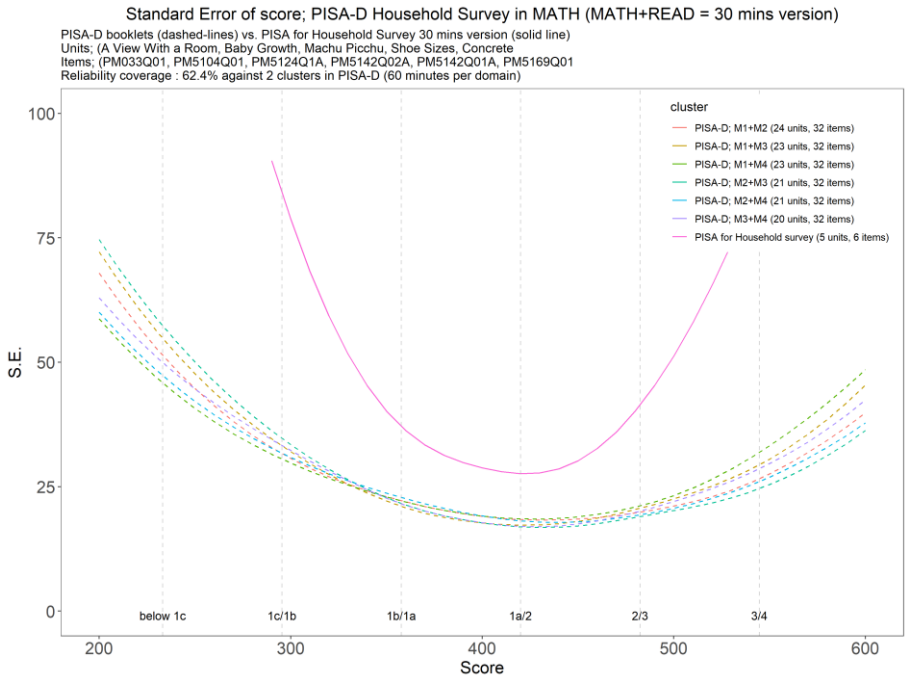


Figure 7 Standard Error of MATH scores; PISA-HSM (30 mins) vs PISA Strand-C

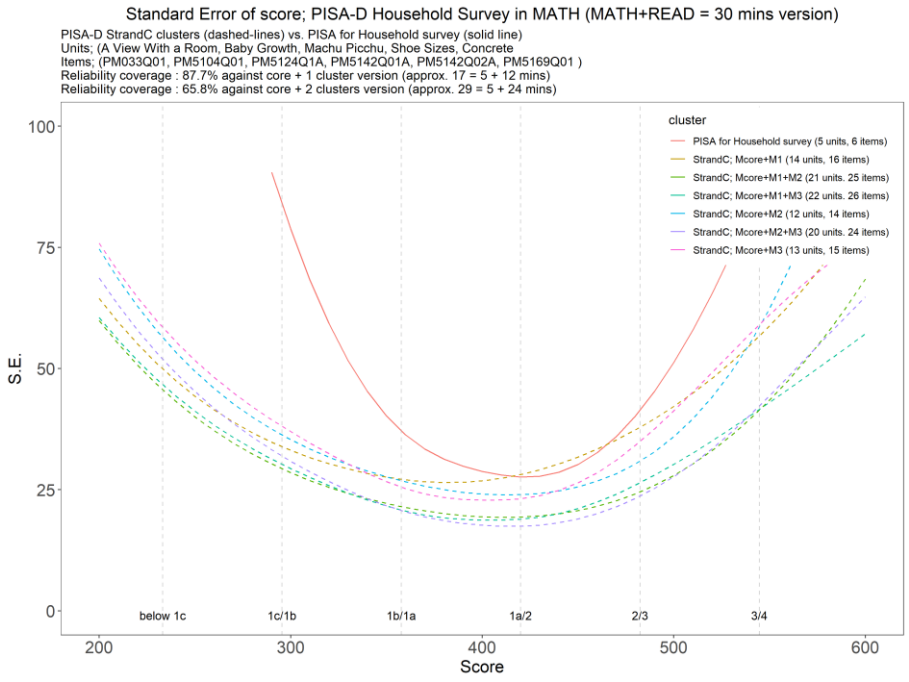


Figure 8 shows the standard error function of the READ item-cluster of PISA-HSM (30 mins) and PISA-D booklets represented in dashed lines. Figure 9 shows the same function but presented together with that of PISA Strand-C.

In READ, the reliability of PISA-HSM is over 65% against PISA-D and 80% against PISA Strand-C. The number of items is 35% comparing to PISA-D, and about 70% comparing to PISA Strand-C. The READ cluster of PISA-HSM (30 mins) is efficient in terms of the balance of reliability and the number of items.

Figure 8 Standard Error of READ scores; PISA-HSM (30 mins version) vs PISA-D

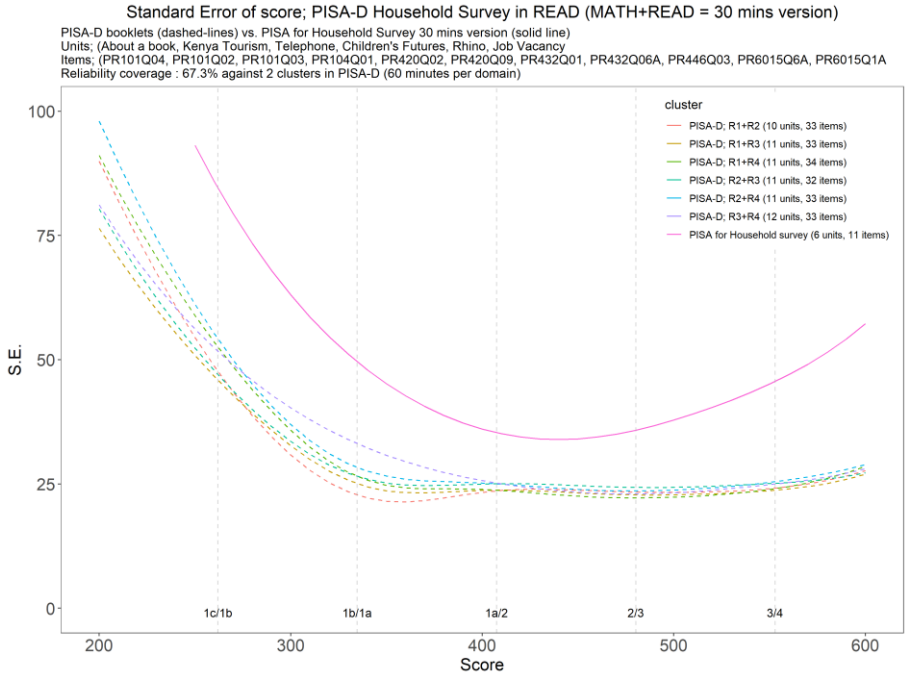
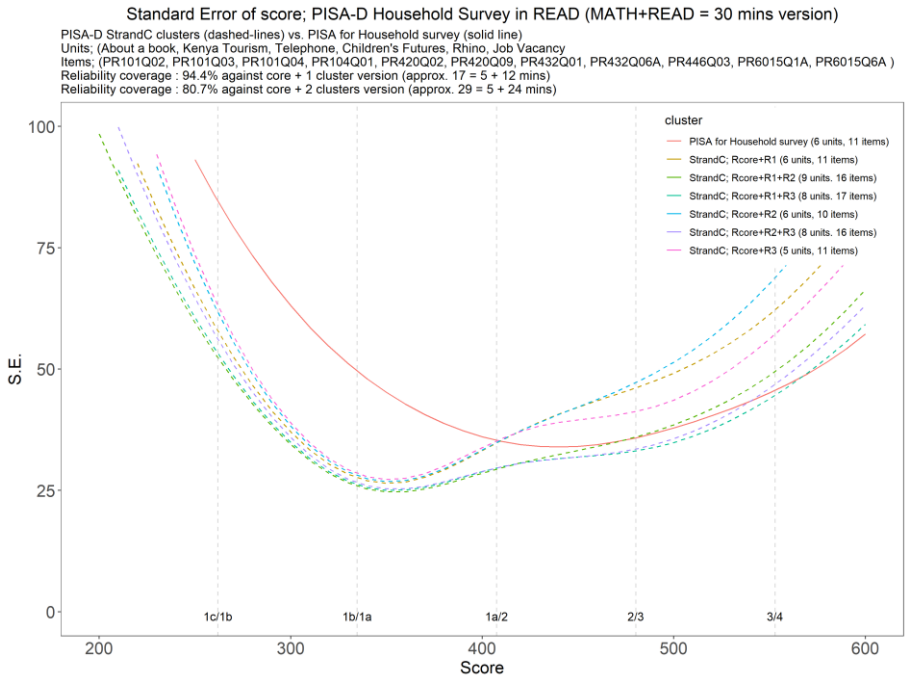


Figure 9 Standard Error of READ scores; PISA-HSM (30 mins version) vs PISA Strand-C



4.3. PISA-HSM (45 minutes version)

Item cluster of PISA-HSM (45 minutes version)

The selected items for the 45 minutes version of PISA-HSM are shown in Table 11 and Table 12. The selected 8 MATH items (7 units) cover 68.9% of reliability against the PISA-D and 96.9% against the PISA Strand-C. For READ, 17 items (9 units) were selected in PISA-HSM (45 mins). It covers 79.7% of reliability against the PISA-D booklets and 95.6% against the PISA Strand-C.

Table 11 Selected MATH items for PISA-HSM (45 minutes version)

Item	Unit	Category 1	Category 2	Source	Format
PM033Q01	A View With a Room	Space and Shape	Interpret	PISA 2015 Trend	MCQ
PM447Q01	Tile Arrangement	Space and Shape	Employ	PISA 2015 Trend	MCQ
PM5104Q01	Baby Growth	Uncertainty and data	Formulate	PISA for Schools	MCQ
PM5124Q1A	Concrete Path	Change and relationships	Formulate	PISA for Schools	KW
PM5142Q01A	Machu Picchu	Quantity	Interpret	PISA for Schools	KW
PM5142Q02A	Machu Picchu	Quantity	Employ	PISA for Schools	KW
PM5169Q01*	Shoe Sizes	Change and relationships	Employ	PISA for Schools	KW
PM5188Q01A	Car Speed	Uncertainty and data	Employ	PISA for Schools	KW

Note: Items highlighted in yellow are still in use in PISA 2022 reading (CBA). Items marked by an asterisk were *not* used in Strand C. All items are still in use in the PISA 2022 PBA assessment.

Table 12 Selected READ items for PISA-HSM (45 minutes version)

Item	Unit	Category 1	Category 2	Source	Format
PR055Q01	Drugged Spiders	Integrate and interpret	Public	PISA 2015 Trend	MCQ
PR101Q02	Rhino	Integrate and interpret	Public	PISA 2015 Trend	MCQ
PR101Q03	Rhino	Reflect and evaluate	Public	PISA 2015 Trend	MCQ
PR101Q04	Rhino	Integrate and interpret	Public	PISA 2015 Trend	MCQ
PR104Q01*	Telephone	Access and retrieve	Public	PISA 2015 Trend	KW
PR220Q02B*	South Pole	Integrate and interpret	Educational	PISA 2015 Trend	MCQ
PR220Q05	South Pole	Integrate and interpret	Educational	PISA 2015 Trend	MCQ
PR220Q06	South Pole	Integrate and interpret	Educational	PISA 2015 Trend	MCQ
PR420Q02	Children's Futures	Access and retrieve	Educational	PISA 2015 Trend	KW
PR420Q09	Children's Futures	Access and retrieve	Educational	PISA 2015 Trend	KW
PR432Q01*	About a book	Integrate and interpret	Personal	PISA 2015 Trend	KW
PR432Q06A	About a book	Integrate and interpret	Personal	PISA 2015 Trend	X
PR446Q03	Job Vacancy	Access and retrieve	Occupational	PISA 2015 Trend	KW
PR460Q05*	Gulf of Mexico	Access and retrieve	Educational	PISA 2015 Trend	MCQ
PR460Q06*	Gulf of Mexico	Integrate and interpret	Educational	PISA 2015 Trend	MCQ
PR6015Q1A	Kenya Tourism	Reflect and evaluate	Educational	PISA for Schools	MCQ
PR6015Q6A	Kenya Tourism	Access and retrieve	Educational	PISA for Schools	MCQ

Note: Items highlighted in yellow are still in use in PISA 2022 reading (CBA). Items marked by an asterisk were *not* used in Strand C. All items are still in use in the PISA 2022 PBA assessment.

Booklet design of PISA-HSM (45 mins version)

Two types of booklet designs were considered in PISA-HSM (45 mins). Table 13 shows item-clusters of the test. The READ items were split into R1 cluster and R2 cluster, while the MATH items were all assigned to M1 cluster. Table 14 shows two booklet patterns; Pattern A and Pattern B. In Pattern A, two booklets were assembled, as shown in the table. M1 was placed at the first block in both booklets. In Pattern B, two booklets were assembled in which M1 was fixed at the second block. Both patterns try to avoid the risk of not completing the MATH items.

Table 13 Item clusters of PISA-HSM (45 minutes version)

#	M1		R1		R2	
1	PM033Q01	A View With a Room	PR055Q01	Drugged Spiders	PR420Q02	Children's Futures
2	PM447Q01	Tile Arrangement	PR101Q02	Rhino	PR420Q09	Children's Futures
3	PM5104Q01	Baby Growth	PR101Q03	Rhino	PR432Q01	About a book
4	PM5124Q1A	Concrete Path	PR101Q04	Rhino	PR432Q06A	About a book
5	PM5142Q01A	Machu Picchu	PR104Q01	Telephone	PR446Q03	Job Vacancy
6	PM5142Q02A	Machu Picchu	PR220Q02B	South Pole	PR460Q05	Gulf of Mexico
7	PM5169Q01	Shoe Sizes	PR220Q05	South Pole	PR460Q06	Gulf of Mexico
8	PM5188Q01A	Car Speed	PR220Q06	South Pole	PR6015Q1A	Kenya Tourism
9					PR6015Q6A	Kenya Tourism

Table 14 Booklet design of PISA-D HSM (45 minutes version)

	Pattern A		Pattern B	
	Booklet #1	**Booklet #2**	**Booklet #1**	**Booklet #2**
1	M1	M1	R1	R2
2	R1	R2	M1	M1
3	R2	R1	R2	R1

Reliability of PISA-HSM (45 mins version)

The standard error function of the MATH in PISA-HSM (45 mins) is shown in Figure 10. For comparisons, the standard error functions of PISA-D booklets are shown in dashed lines. Figure 11 shows the standard error function together with that of PISA Strand-C. The figures show that the PISA-HSM (45 mins) focuses on proficiency levels 1a/2. In this version, only 25% of the number of items comparing to PISA-D is assigned and keeps about 70% of the reliability. It keeps approximately 95% of the reliability against the PISA Strand-C only with about 50% of items.

Figure 10 Standard Error of MATH scores; PISA-HSM (45 mins version) vs PISA-D

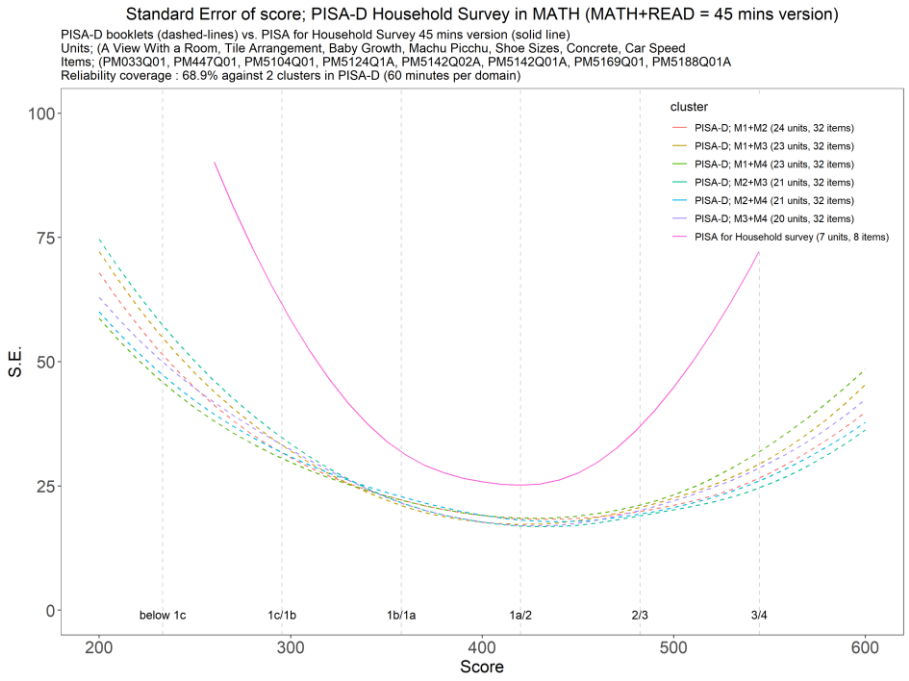


Figure 11 Standard Error of MATH scores; PISA-HSM (45 mins version) vs PISA Strand C

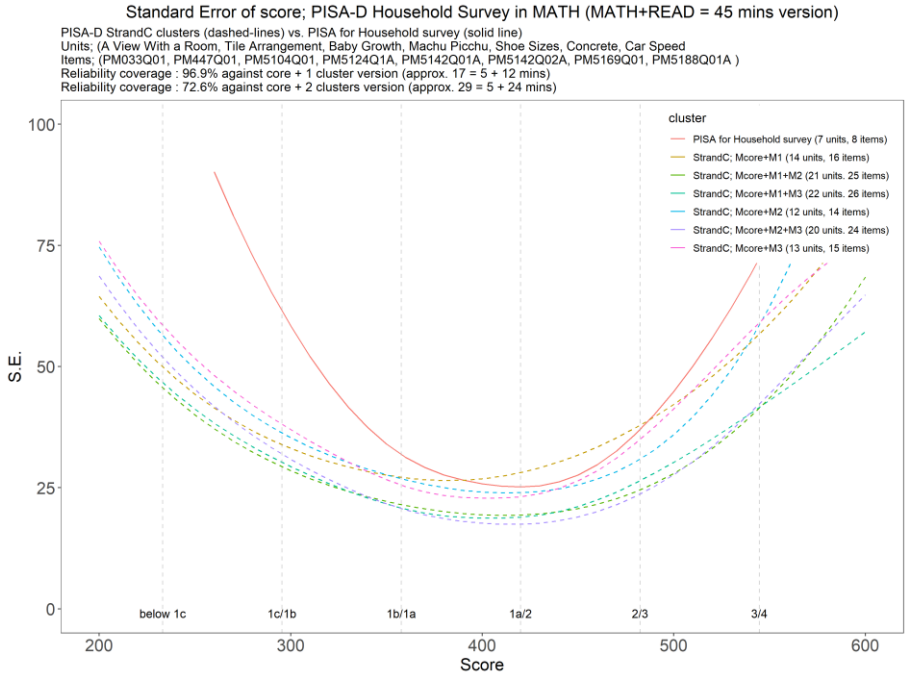


Figure 12 shows the standard error function of the READ cluster of PISA-HSM (45 mins), which also shows that of PISA-D booklets in dashed lines. Figure 13 shows the same function, but with that of PISA Strand-C READ booklets in dashed lines. The figures show that the PISA-HSM READ cluster covers a broader range of proficiency

levels than the MATH cluster. In READ, the reliability of the PISA-HSM (45 mins) is approximately 80% comparing to PISA-D, and over 95% reliability comparing to PISA Strand-C READ booklets (core + 2 clusters). The number of items in PISA-HSM (45 mins) is about 50% of PISA-D, and the same with that of PISA Strand-C READ booklets (core + 2 clusters).

Figure 12 Standard Error of READ scores; PISA-HSM (45 mins version) vs PISA-D

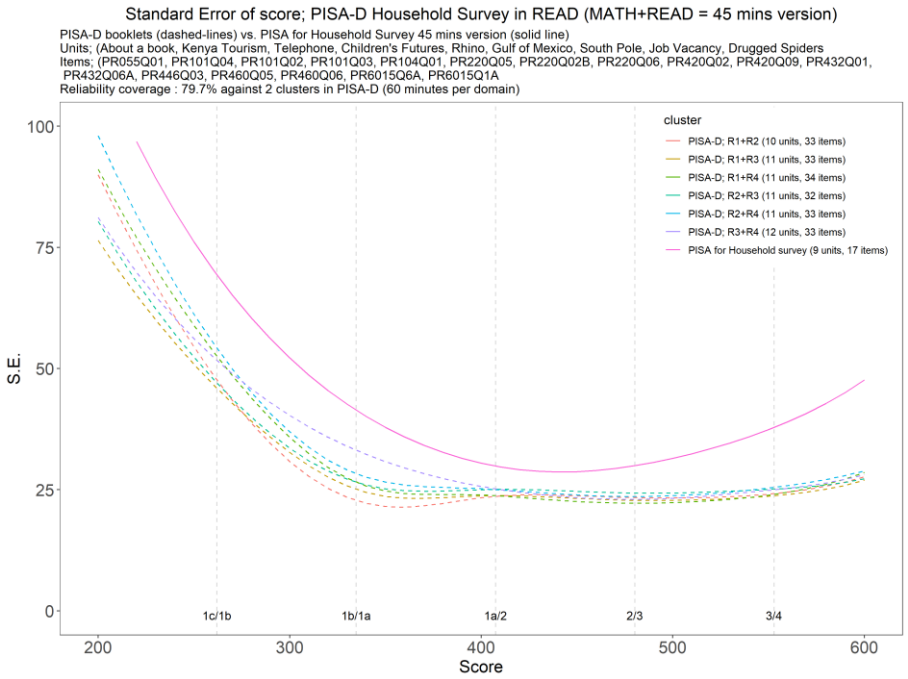
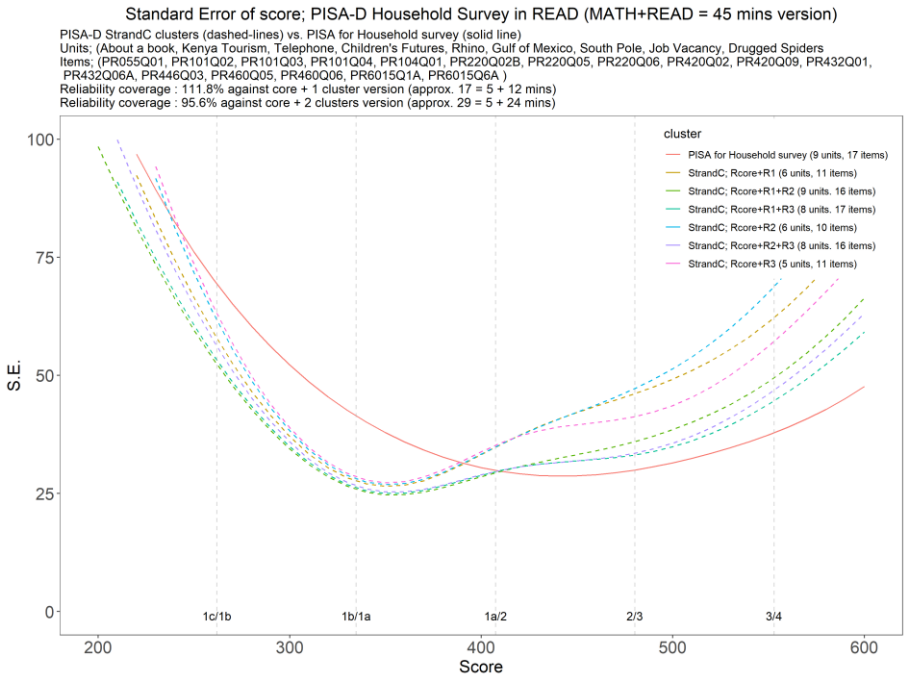


Figure 13 Standard Error of READ scores; PISA-HSM (45 mins version) vs PISA Strand-C



4.4. Comparisons among PISA-D family

Table 15 represents the comparisons among different versions of PISA-D tests. The table shows that the numbers of items per minute in PISA Strand-C are much more than PISA-D and PISA-HSM. PISA-HSM (45 mins) keeps the same level of reliability as PISA Strand-C for both MATH and READ, but the number of items in MATH is much less than that of PISA Strand-C. The PISA-HSM (45 mins) is succeeded in keeping the reliability of the scores with a fewer number of items.

PISA-HSM 30 minutes version is much more efficient than the 45 minutes version. It keeps 80-90% of reliability against PISA Strand-C with much fewer MATH items comparing to it.

The standard error of a student score around the target levels is an acceptable level (both the 30 minutes version and the 45 minutes version). The standard error of a group level average score is a function of $\sqrt{N}$, therefore it will be sufficiently small.

Table 15 Comparisons among test versions of PISA-D

	PISA-D		PISA Strand-C		PISA-HSM (30 mins)		PISA-HSM (45 mins)	
	MATH	READ	MATH	READ	MATH	READ	MATH	READ
Test length (mins)	60	60	45		30 (MATH 10 + READ 20)		45 (MATH 15 + READ 30)	
N of units	20-24	10-12	12-22	5-9	5	6	7	9
N of items	32	32-34	14-26	11-17	6	11	8	17
SE around target level			19.1-25.5	29.0-34.0	29.1	36.0	26.4	30.4
Coverage of reliability against PISA-D					62.4%	67.3%	68.9%	79.7%
Coverage of reliability against Strand-C					87.7%	80.7%	96.9%	95.6%