



12

Scaling Outcomes

| | |
|---|-----|
| International characteristics of the item pool | 188 |
| Scaling outcomes | 195 |
| Test length analysis | 199 |
| Booklet effects | 203 |
| Observations concerning the construction of the PISA overall literacy scales | 213 |
| Transforming the plausible values to PISA scales | 229 |
| Link error | 230 |

This chapter describes the application of Item Response Theory (IRT) scaling and plausible value methodology to the PISA 2009 assessment data.

INTERNATIONAL CHARACTERISTICS OF THE ITEM POOL

When main study data were received from each participating country, they were first verified and cleaned using the procedures outlined in Chapter 10. Files containing the achievement data were prepared and national-level Rasch and traditional test analyses were undertaken. The results of these analyses were included in the reports that were returned to each participant (see Chapter 9).

After processing at the national level, a set of international-level analyses was undertaken. Some involved summarising national analyses, while others required an analysis of the international data set.

The final international cognitive data set (that is, the data set of coded achievement booklet responses – available as *INT_cogn09_TD_Dec10.txt*) consisted of 475 460 students from 65 participating countries. Table 12.1 shows the total number of students included in the *PISA 2009 Database*, broken down by participating country and test booklet. Countries that implemented the easier (see Chapter 2) set of booklets are marked with an * in this table.

Nineteen countries participated in PISA 2009 digital reading assessment (DRA). The number of the cases included in DRA cognitive data set is the same as in international cognitive data set for all participating countries except for Colombia and Spain, which have chosen to have schools sub sampled for DRA from their large national school sample (see Chapter 4 for details of DRA sampling).

Proficiency estimates were imputed for the students that were not sampled for DRA. The final international DRA cognitive data file (available as *ERA_cogn09_TD_Jun11.txt*) contains 107 394 students. Table 12.2 shows the total number of students included in the *PISA DRA 2009 Database*, broken down by participating country and DRA test form. For the students that were not sampled for DRA, the test form code is 7.

[Part 1/2]
Table 12.1 Number of sampled students by country and booklet

| | Booklets | | | | | | | | | | | | | | Total | |
|-----------------|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----|--------|--|
| | 1(21) | 2(22) | 3(23) | 4(24) | 5(25) | 6(26) | 7(27) | 8 | 9 | 10 | 11 | 12 | 13 | UH | | |
| OECD | | | | | | | | | | | | | | | | |
| Australia | 1 094 | 1 107 | 1 079 | 1 081 | 1 096 | 1 081 | 1 079 | 1 114 | 1 133 | 1 094 | 1 112 | 1 092 | 1 089 | | 14 251 | |
| Austria | 496 | 503 | 499 | 515 | 519 | 506 | 498 | 491 | 489 | 487 | 490 | 498 | 489 | 110 | 6 590 | |
| Belgium | 646 | 615 | 622 | 644 | 653 | 643 | 647 | 634 | 625 | 611 | 618 | 631 | 621 | 291 | 8 501 | |
| Canada | 1 767 | 1 788 | 1 786 | 1 793 | 1 793 | 1 799 | 1 792 | 1 746 | 1 814 | 1 782 | 1 758 | 1 810 | 1 779 | | 23 207 | |
| Chile* | 444 | 422 | 425 | 434 | 468 | 432 | 437 | 440 | 430 | 444 | 417 | 434 | 442 | | 5 669 | |
| Czech Republic | 459 | 462 | 436 | 463 | 443 | 451 | 432 | 430 | 443 | 447 | 455 | 461 | 460 | 222 | 6 064 | |
| Denmark | 445 | 443 | 465 | 468 | 465 | 463 | 454 | 439 | 459 | 455 | 447 | 460 | 461 | | 5 924 | |
| Estonia | 367 | 354 | 357 | 357 | 352 | 366 | 363 | 360 | 361 | 379 | 369 | 372 | 370 | | 4 727 | |
| Finland | 454 | 446 | 453 | 449 | 446 | 447 | 438 | 440 | 438 | 455 | 444 | 454 | 446 | | 5 810 | |
| France | 332 | 320 | 334 | 314 | 319 | 335 | 333 | 312 | 344 | 334 | 339 | 333 | 349 | | 4 298 | |
| Germany | 382 | 369 | 362 | 362 | 379 | 371 | 360 | 370 | 367 | 379 | 366 | 363 | 370 | 179 | 4 979 | |
| Greece | 389 | 382 | 385 | 381 | 376 | 369 | 381 | 381 | 386 | 385 | 380 | 385 | 389 | | 4 969 | |
| Hungary | 352 | 354 | 349 | 359 | 349 | 361 | 357 | 350 | 355 | 357 | 349 | 359 | 354 | | 4 605 | |
| Iceland | 280 | 275 | 282 | 279 | 279 | 286 | 278 | 281 | 282 | 280 | 286 | 274 | 284 | | 3 646 | |
| Ireland | 300 | 305 | 300 | 282 | 299 | 294 | 320 | 308 | 308 | 319 | 299 | 296 | 307 | | 3 937 | |
| Israel | 810 | 411 | 429 | 406 | 407 | 417 | 417 | 420 | 416 | 408 | 406 | 416 | 398 | | 5 761 | |
| Italy | 2 366 | 2 359 | 2 383 | 2 386 | 2 401 | 2 416 | 2 389 | 2 370 | 2 367 | 2 356 | 2 378 | 2 369 | 2 365 | | 30 905 | |
| Japan | 457 | 468 | 470 | 465 | 466 | 467 | 465 | 470 | 470 | 464 | 472 | 478 | 476 | | 6 088 | |
| Korea | 374 | 382 | 377 | 387 | 386 | 399 | 392 | 393 | 394 | 381 | 379 | 371 | 374 | | 4 989 | |
| Luxembourg | 352 | 359 | 355 | 357 | 360 | 355 | 357 | 361 | 351 | 349 | 360 | 352 | 354 | | 4 622 | |
| Mexico* | 2 973 | 2 953 | 2 959 | 2 959 | 2 942 | 2 973 | 2 933 | 2 935 | 2 908 | 2 910 | 2 948 | 2 917 | 2 940 | | 38 250 | |
| Netherlands | 359 | 355 | 376 | 357 | 362 | 368 | 362 | 357 | 356 | 355 | 347 | 348 | 361 | 97 | 4 760 | |
| New Zealand | 364 | 357 | 350 | 351 | 350 | 354 | 364 | 361 | 361 | 344 | 355 | 370 | 362 | | 4 643 | |
| Norway | 354 | 352 | 355 | 360 | 372 | 369 | 352 | 365 | 355 | 356 | 361 | 361 | 348 | | 4 660 | |
| Poland | 381 | 394 | 382 | 368 | 381 | 372 | 370 | 370 | 384 | 375 | 386 | 372 | 382 | | 4 917 | |
| Portugal | 496 | 451 | 494 | 487 | 482 | 484 | 494 | 490 | 489 | 466 | 500 | 480 | 485 | | 6 298 | |
| Slovak Republic | 343 | 320 | 338 | 341 | 339 | 341 | 355 | 353 | 372 | 362 | 358 | 355 | 348 | 30 | 4 555 | |
| Slovenia | 460 | 456 | 459 | 457 | 455 | 460 | 471 | 471 | 462 | 462 | 457 | 469 | 454 | 162 | 6 155 | |
| Spain | 1 983 | 1 952 | 2 004 | 2 033 | 1 995 | 1 993 | 1 981 | 2 008 | 1 989 | 2 023 | 2 002 | 1 965 | 1 959 | | 25 887 | |
| Sweden | 349 | 360 | 354 | 351 | 349 | 357 | 339 | 344 | 348 | 347 | 351 | 361 | 357 | | 4 567 | |
| Switzerland | 917 | 897 | 882 | 936 | 930 | 865 | 915 | 906 | 905 | 922 | 881 | 908 | 948 | | 11 812 | |
| Turkey | 388 | 386 | 378 | 382 | 373 | 385 | 380 | 392 | 390 | 385 | 389 | 383 | 385 | | 4 996 | |
| United Kingdom | 939 | 944 | 932 | 921 | 927 | 933 | 916 | 926 | 957 | 934 | 957 | 951 | 942 | | 12 179 | |
| United States | 406 | 400 | 409 | 400 | 402 | 396 | 398 | 402 | 413 | 406 | 407 | 398 | 396 | | 5 233 | |

*These countries opted for the easier booklets.



[Part 2/2]
Table 12.1 Number of sampled students by country and booklet

| | Booklets | | | | | | | | | | | | | | UH | Total |
|----------------------|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--|----|--------|
| | 1(21) | 2(22) | 3(23) | 4(24) | 5(25) | 6(26) | 7(27) | 8 | 9 | 10 | 11 | 12 | 13 | | | |
| <i>Partners</i> | | | | | | | | | | | | | | | | |
| Albania* | 352 | 348 | 352 | 340 | 351 | 358 | 374 | 367 | 362 | 353 | 345 | 343 | 351 | | | 4 596 |
| Argentina* | 368 | 374 | 367 | 361 | 344 | 370 | 369 | 362 | 369 | 386 | 358 | 366 | 380 | | | 4 774 |
| Azerbaijan* | 354 | 359 | 362 | 367 | 368 | 370 | 375 | 368 | 364 | 359 | 346 | 354 | 345 | | | 4 691 |
| Brazil* | 1 547 | 1 576 | 1 561 | 1 614 | 1 523 | 1 538 | 1 548 | 1 537 | 1 535 | 1 527 | 1 536 | 1 529 | 1 556 | | | 20 127 |
| Bulgaria* | 350 | 350 | 354 | 357 | 349 | 344 | 351 | 347 | 339 | 338 | 337 | 340 | 351 | | | 4 507 |
| Colombia* | 613 | 611 | 602 | 625 | 604 | 600 | 597 | 592 | 612 | 608 | 627 | 625 | 605 | | | 7 921 |
| Croatia | 374 | 368 | 377 | 383 | 386 | 389 | 388 | 387 | 385 | 400 | 397 | 382 | 378 | | | 4 994 |
| Dubai (UAE)* | 411 | 443 | 441 | 444 | 431 | 429 | 438 | 441 | 436 | 432 | 430 | 423 | 421 | | | 5 620 |
| Hong Kong-China | 369 | 374 | 376 | 379 | 380 | 372 | 373 | 363 | 364 | 374 | 367 | 373 | 373 | | | 4 837 |
| Indonesia | 390 | 387 | 382 | 393 | 394 | 391 | 396 | 401 | 398 | 403 | 399 | 402 | 400 | | | 5 136 |
| Jordan* | 505 | 512 | 498 | 493 | 503 | 490 | 491 | 491 | 495 | 491 | 496 | 509 | 512 | | | 6 486 |
| Kazakhstan* | 413 | 409 | 418 | 419 | 415 | 421 | 419 | 427 | 417 | 419 | 406 | 406 | 423 | | | 5 412 |
| Kyrgyzstan* | 397 | 390 | 390 | 381 | 377 | 373 | 379 | 382 | 386 | 377 | 386 | 386 | 382 | | | 4 986 |
| Latvia | 355 | 351 | 354 | 343 | 358 | 342 | 340 | 351 | 340 | 342 | 350 | 333 | 343 | | | 4 502 |
| Liechtenstein | 23 | 24 | 23 | 33 | 25 | 14 | 28 | 23 | 27 | 23 | 27 | 28 | 31 | | | 329 |
| Lithuania | 363 | 345 | 356 | 338 | 354 | 336 | 352 | 338 | 343 | 344 | 351 | 350 | 358 | | | 4 528 |
| Macao-China | 457 | 460 | 456 | 459 | 457 | 459 | 457 | 457 | 459 | 457 | 455 | 457 | 462 | | | 5 952 |
| Montenegro | 367 | 369 | 372 | 360 | 373 | 383 | 375 | 376 | 376 | 368 | 371 | 379 | 356 | | | 4 825 |
| Panama* | 299 | 308 | 312 | 297 | 307 | 303 | 312 | 312 | 302 | 306 | 302 | 302 | 307 | | | 3 969 |
| Peru* | 465 | 472 | 458 | 474 | 459 | 456 | 443 | 449 | 459 | 454 | 465 | 470 | 461 | | | 5 985 |
| Qatar* | 696 | 681 | 699 | 706 | 713 | 707 | 701 | 696 | 697 | 699 | 702 | 688 | 693 | | | 9 078 |
| Romania* | 368 | 359 | 359 | 355 | 358 | 372 | 374 | 378 | 373 | 374 | 372 | 364 | 370 | | | 4 776 |
| Russian Federation | 406 | 414 | 415 | 409 | 409 | 410 | 409 | 402 | 398 | 403 | 412 | 413 | 408 | | | 5 308 |
| Serbia* | 417 | 426 | 434 | 434 | 439 | 426 | 429 | 430 | 416 | 422 | 417 | 415 | 418 | | | 5 523 |
| Shanghai-China | 400 | 398 | 397 | 388 | 386 | 392 | 387 | 391 | 385 | 394 | 398 | 404 | 395 | | | 5 115 |
| Singapore | 412 | 405 | 402 | 408 | 408 | 416 | 413 | 410 | 404 | 401 | 394 | 406 | 404 | | | 5 283 |
| Chinese Taipei | 445 | 445 | 447 | 452 | 451 | 438 | 452 | 452 | 448 | 441 | 450 | 449 | 461 | | | 5 831 |
| Thailand | 489 | 486 | 475 | 478 | 476 | 476 | 480 | 483 | 473 | 471 | 473 | 478 | 487 | | | 6 225 |
| Trinidad and Tobago* | 369 | 351 | 355 | 364 | 359 | 368 | 358 | 366 | 380 | 377 | 379 | 384 | 368 | | | 4 778 |
| Tunisia* | 381 | 377 | 394 | 370 | 382 | 368 | 376 | 382 | 384 | 384 | 389 | 382 | 386 | | | 4 955 |
| Uruguay* | 464 | 455 | 452 | 454 | 456 | 460 | 467 | 449 | 458 | 453 | 467 | 466 | 456 | | | 5 957 |

Table 12.2 Number of sampled students by country and DRA test form code

| | Booklets | | | | | | Total sampled students | Total not-sampled students |
|-----------------|----------|-----|-----|-----|-----|-----|------------------------|----------------------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | | |
| <i>OECD</i> | | | | | | | | |
| Australia | 496 | 520 | 505 | 495 | 483 | 494 | 2 993 | 11 258 |
| Austria | 454 | 450 | 437 | 417 | 426 | 447 | 2 631 | 3 959 |
| Belgium | 485 | 474 | 448 | 457 | 475 | 469 | 2 808 | 5 693 |
| Chile | 297 | 288 | 278 | 274 | 280 | 287 | 1 704 | 3 965 |
| Denmark | 220 | 208 | 198 | 210 | 208 | 226 | 1 270 | 4 654 |
| France | 216 | 203 | 221 | 213 | 228 | 224 | 1 305 | 2 993 |
| Hungary | 311 | 298 | 286 | 302 | 298 | 298 | 1 793 | 2 812 |
| Ireland | 249 | 233 | 239 | 219 | 236 | 233 | 1 409 | 2 528 |
| Israel | 155 | 159 | 163 | 165 | 164 | 156 | 962 | 2 684 |
| Japan | 582 | 575 | 570 | 577 | 575 | 550 | 3 429 | 2 659 |
| Korea | 255 | 247 | 249 | 239 | 237 | 250 | 1 477 | 3 512 |
| New Zealand | 296 | 292 | 301 | 286 | 286 | 294 | 1 755 | 2 888 |
| Norway | 338 | 329 | 310 | 326 | 340 | 331 | 1 974 | 2 686 |
| Poland | 350 | 347 | 321 | 326 | 314 | 330 | 1 988 | 2 929 |
| Spain | 283 | 277 | 269 | 291 | 285 | 284 | 1 689 | 3 059 |
| Sweden | 336 | 308 | 313 | 323 | 328 | 313 | 1 921 | 2 646 |
| <i>Partners</i> | | | | | | | | |
| Colombia | 496 | 520 | 505 | 495 | 483 | 494 | 2 993 | 11 258 |
| Hong Kong-China | 454 | 450 | 437 | 417 | 426 | 447 | 2 631 | 3 959 |
| Macao-China | 485 | 474 | 448 | 457 | 475 | 469 | 2 808 | 5 693 |

Test targeting

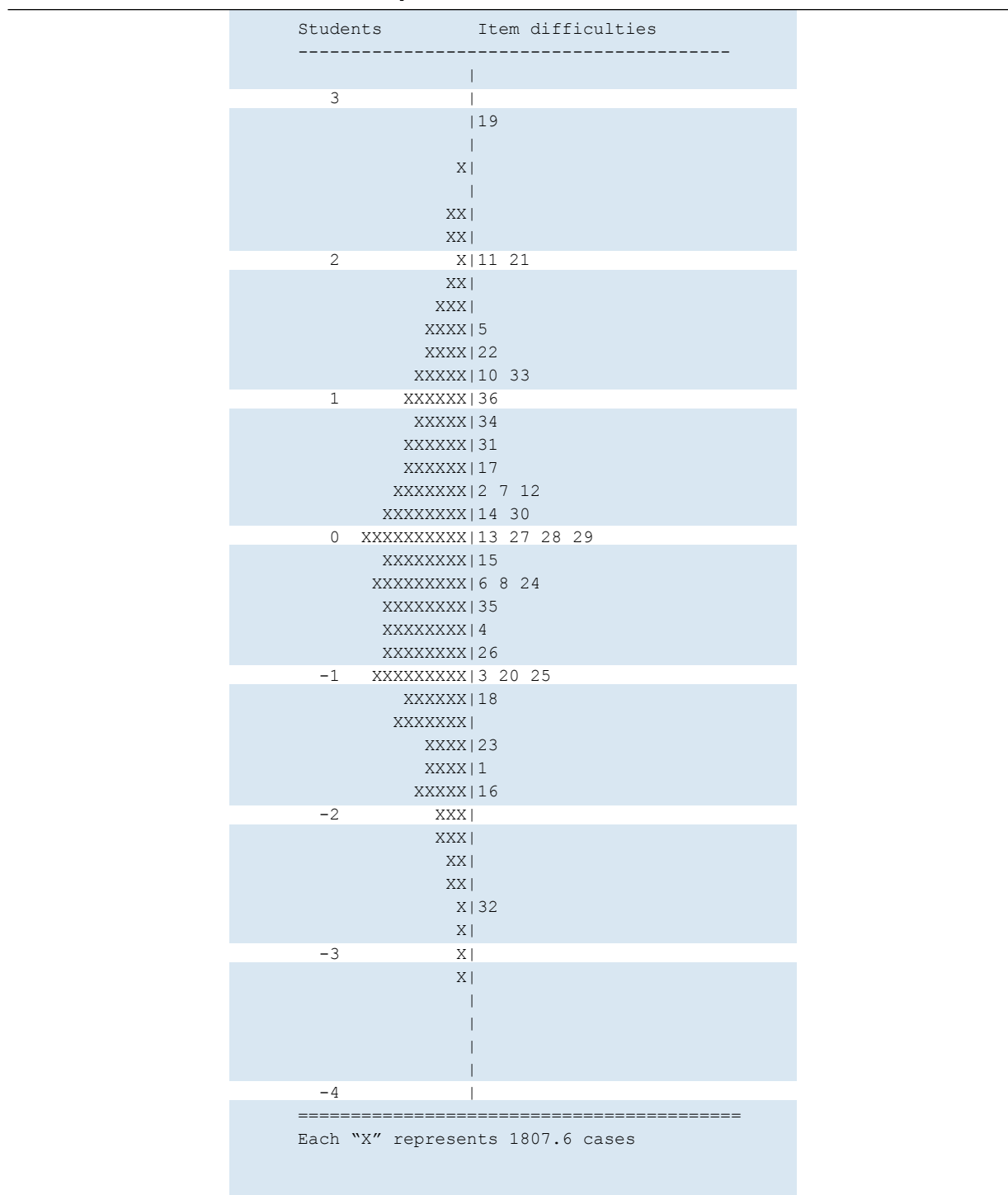
Each of the domains was separately scaled to examine the targeting of the tests. Figures 12.1 to 12.4 show the match between the international (OECD countries only) item difficulty distribution and the distribution of OECD's student achievement for each of reading, mathematics, science and DRA respectively. The figures consist of two panels. The first panel (students) shows the distribution of students' Rasch-scaled achievement estimates. Students at the top end of this distribution have higher proficiency estimates than the students at the lower end of the distribution. The second panel (item difficulties) shows the distribution of Rasch-estimated item difficulties.

Test is well targeted if the average of item difficulties is about the same as the average of the students' abilities and the item difficulties are evenly spread across the ability distribution.

In each of the Figures 12.1 to 12.4, the student proficiency distribution shown by Xs^1 is well matched to the item difficulty distribution. The figures are constructed so that when a student and an item are located at the same location on the scale then the student has a 50% chance of responding correctly to the item.

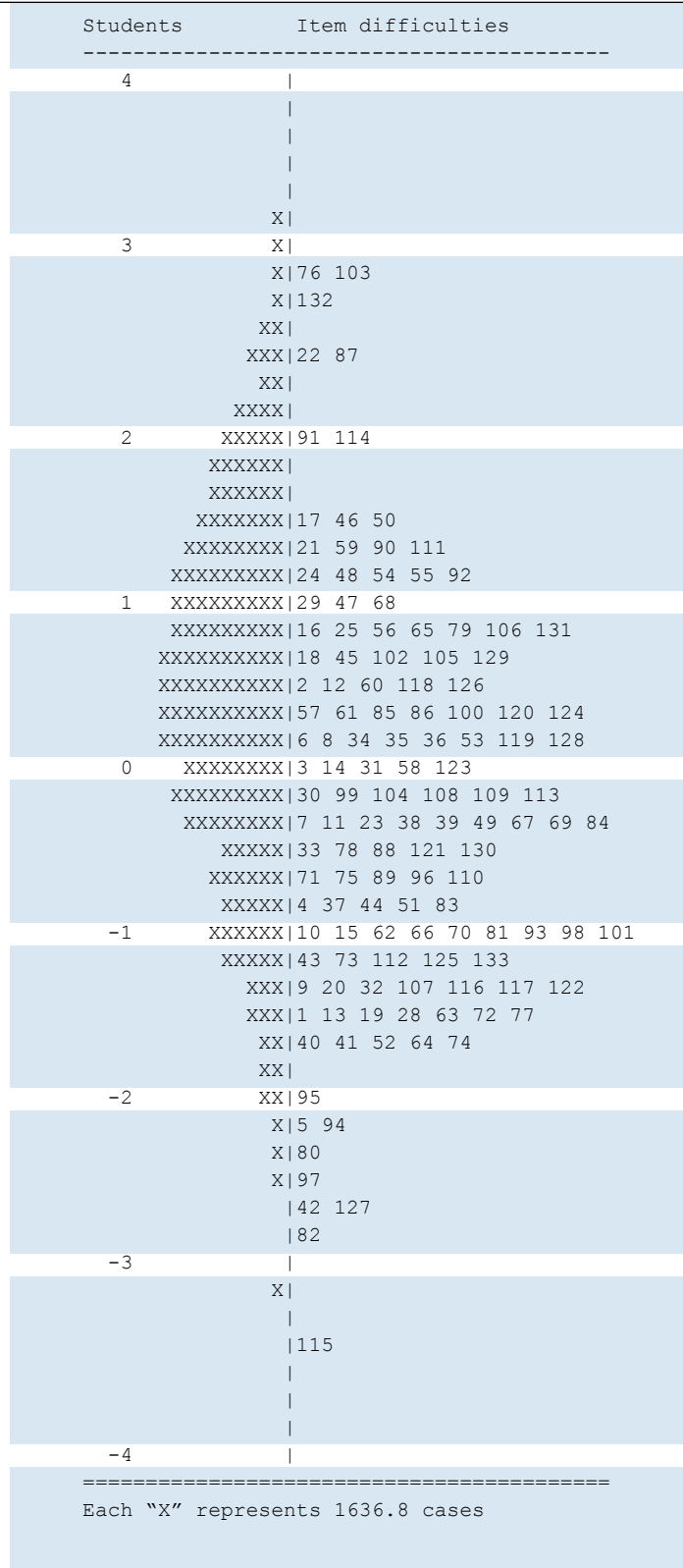
■ Figure 12.1 ■

Item plot for mathematics items

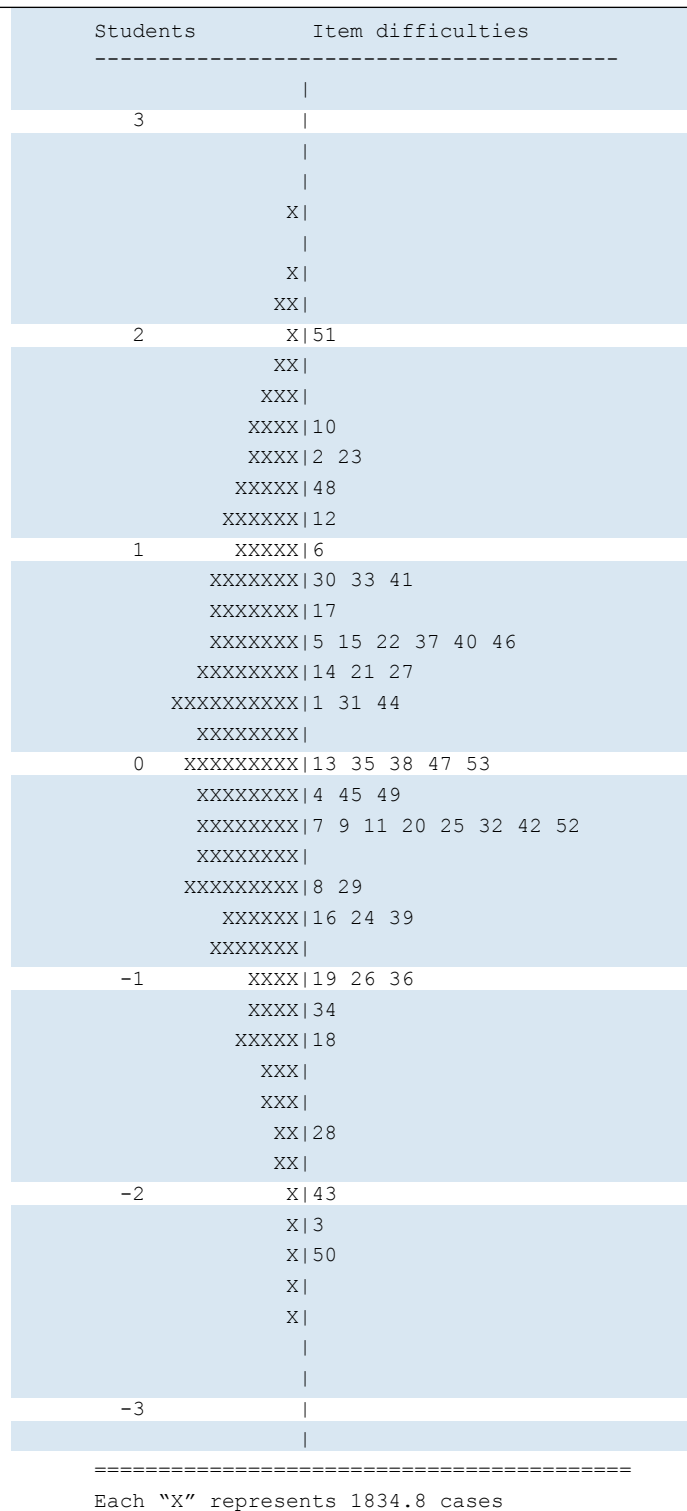




■ Figure 12.2 ■
Item plot for reading items

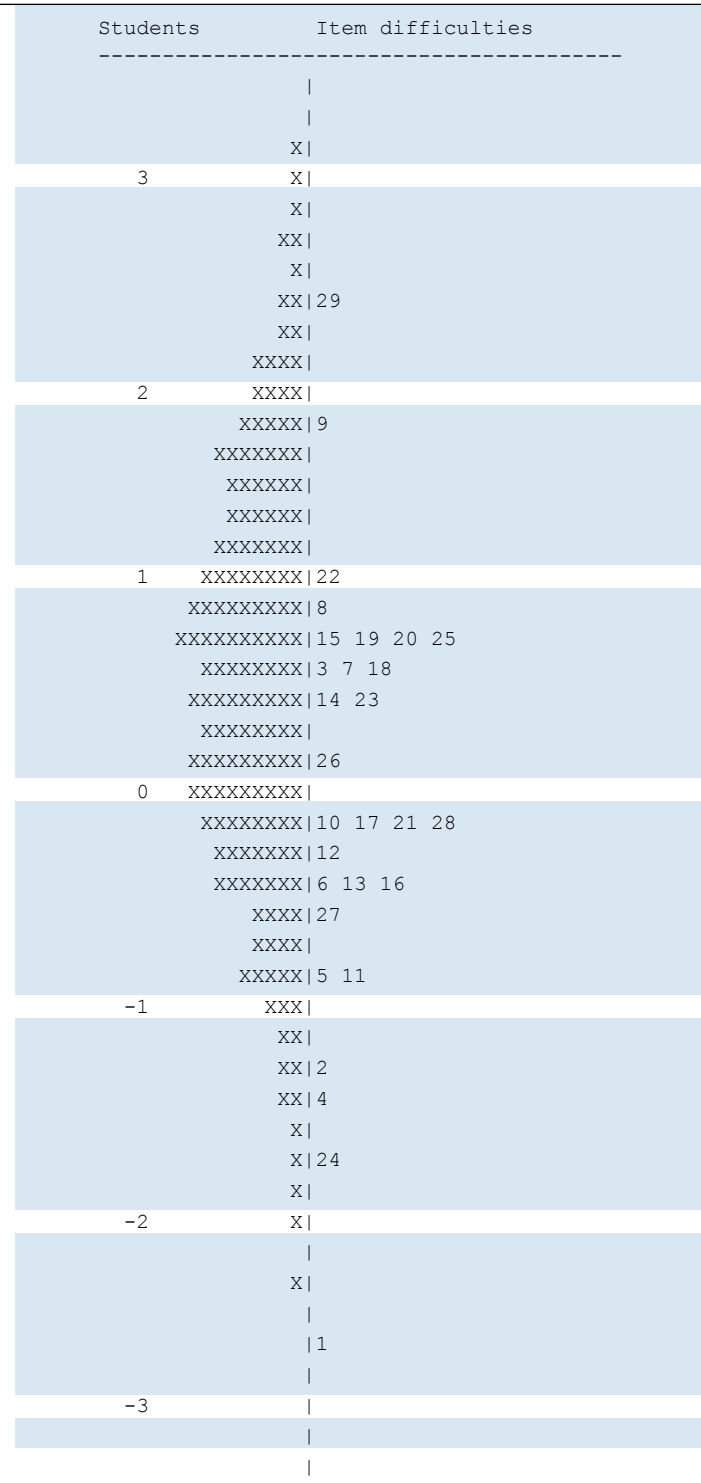


■ Figure 12.3 ■
Item plot for science items





■ Figure 12.4 ■
Item plot for DRA items



=====
 Each "X" represents 636.4 cases

Test reliability and measurement error design effect

A second test characteristic that is of importance is the test reliability, or equivalently the measurement error design effect (Adams, 2005). Table 12.3 shows the reliability for each of the three overall scales (mathematical literacy, reading literacy and scientific literacy) and for the DRA scale before conditioning and based upon four separate unidimensional scalings, using plausible values (PV) and using Weighted Likelihood Estimates (WLE).

The WLE-based estimates are IRT analogues of traditional estimates of Person separation reliability such as internal consistency. They are estimated for the samples of students that responded to test forms from each of the domains.

The plausible value based estimates, however, use all sampled students and represent the influence of the test design on the uncertainty of estimates of the overall mean. For example the DRA reliability of 0.30 and corresponding design effect of 3.33 means that the error variance of the estimate of the mean would be increased by a factor of 3.33 because of the use of a sub-sample and seven alternative assessment booklets. These estimates take into account the fact that the sample sizes for each domain are markedly different. The consequence is that the WLE reliabilities for the minor domains are higher than the PV reliabilities because students that were not assessed in mathematics, science or DRA were excluded from the calculation of the WLE reliabilities.

The plausible value based estimates in Table 12.2 are based upon unidimensional scaling, and do not reflect the benefit of the conditioning and the multidimensional scaling that is implemented in PISA. The international reliability for each domain after conditioning and multidimensional scaling is reported in Table 12.9.

Table 12.3 Reliabilities and Measurement Error Design Effect of each of the three overall scales when scaled separately

| Domain | Reliability (WLE) | Measurement Error Design Effect (WLE) | Reliability (PV) | Measurement Error Design Effect (PV) |
|-------------|-------------------|---------------------------------------|------------------|--------------------------------------|
| Mathematics | 0.74 | 1.34 | 0.54 | 1.84 |
| Reading | 0.84 | 1.19 | 0.86 | 1.17 |
| Science | 0.80 | 1.26 | 0.57 | 1.75 |
| DRA | 0.85 | 1.18 | 0.30 | 3.33 |

Domain inter-correlations

Correlations between the ability estimates for individual students in each of the three domains, the latent correlations, as estimated by *ConQuest*[®] (Wu, Adams and Wilson, 1997) are given in Table 12.4. Correlations between four domains for countries that implemented DRA are given in Table 12.5. It is important to note that these latent correlations are unbiased estimates of the true correlation between the underlying latent variables. As such they are not attenuated by the unreliability of the measures and will generally be higher than the typical product moment correlations that have not been disattenuated for unreliability. The results in Table 12.4 are reported for both OECD countries and for all participating countries. The results in Table 12.5 are reported for 19 DRA countries.

Table 12.4 Latent correlation between the three domains

| | Reading r | Science r |
|-------------|--------------|--------------|
| Mathematics | | |
| OECD | 0.82 | 0.88 |
| All | 0.84 | 0.89 |
| Reading | | |
| OECD | | 0.87 |
| All | | 0.87 |

Table 12.5 Latent correlation between the four domains

| | Reading r | Science r | DRA r |
|-------------|--------------|--------------|----------|
| Mathematics | 0.83 | 0.91 | 0.80 |
| Reading | | 0.87 | 0.86 |
| Science | | | 0.82 |



Reading scales

As described in Chapter 9, a five-dimensional model consisting of mathematics, science, and the three reading aspect scales: *access and retrieve*, *integrate and interpret*, *reflect and evaluate* was used. Then a four-dimensional model was estimated consisting of mathematics, science, and the two reading text format scales: *continuous text* and *non-continuous text*. Responses from the mathematics and science domains were included in the scaling model to improve the estimation of posterior distributions of the reading scales. The plausible values for mathematics and science generated using these two models were not included in the international database. The correlations between reading subscales as estimated from these two models are given in Table 12.6 and Table 12.7.

Table 12.6 Latent correlation between the aspect reading scales

| | Integrate and interpret r | Reflect and evaluate r |
|--------------------------------|------------------------------|---------------------------|
| Access and retrieve | | |
| OECD | 0.93 | 0.90 |
| All | 0.96 | 0.93 |
| Integrate and interpret | | |
| OECD | | 0.94 |
| All | | 0.95 |

Table 12.7 Latent correlation between text format reading scales

| | Non-continuous text r |
|------------------------|--------------------------|
| Continuous text | |
| OECD | 0.93 |
| All | 0.95 |

SCALING OUTCOMES

The procedures for the national and international scaling are outlined in Chapter 9 and are not reiterated here.

National item deletions

The items were first scaled by country and their fit was considered at the national level, as was the consistency of the item parameter estimates across countries. Consortium staff then adjudicated items, considering the items' functioning both within and across countries in detail. Those items considered to be dodgy (see Chapter 9) were then reviewed in consultation with National Project Managers (NPMs). The consultations resulted in the deletion of a number of items at the national level.

At the international level, two reading items (*R219Q01E* and *R219Q01T*) and one mathematics item (*M305Q01*) were deleted from scaling. *R219Q01E* and *R219Q01T* were deleted because of data entry errors and *M305Q01* was deleted because instruction to have a rule was not included in the booklets. The nationally deleted items are listed in Table 12.8. All deleted items were recoded as not applicable and were excluded from both international scaling and generating plausible values.

[Part 1/2]

Table 12.8 Items deleted at the national level

| Item | Country |
|----------|---|
| M033Q01 | Hungary (booklet 8), Serbia |
| M155Q01 | Peru, Dubai (UAE) (Arabic-language version) |
| M305Q01 | International Deletion |
| M406Q01 | Israel (booklet 7 of Arabic-language version) |
| M408Q01T | Denmark |
| M442Q02 | Belgium (booklet 5 of Dutch-language version), Spain (Euskara-language version), Poland (booklet 1), Dubai (UAE) (Arabic-language version), Qatar (Arabic-language version) |
| M474Q01 | Hong Kong-China (Cantonese-language version) |
| M571Q01 | Greece (booklet 10) |
| M603Q01T | Belgium (Dutch-language version) |
| M828Q01 | Israel (Hebrew-language version), Dubai (UAE) (Arabic-language version) |
| M828Q03 | Indonesia |
| R055Q01 | Hungary, Serbia (Serbian-language version) |
| R067Q01 | Switzerland (booklet 2 of Italian-language version) |
| R067Q04 | Switzerland (booklet 2 of Italian-language version) |
| R067Q05 | Switzerland (booklet 2 of Italian-language version), Chile |
| R083Q01 | Hungary |
| R101Q05 | Iceland (booklet 5) |
| R102Q04A | Argentina |
| R102Q05 | Argentina, Hong Kong-China (Cantonese-language version), Macao-China (Cantonese-language version), Mexico (booklet 13), Montenegro (Serbian/variant of Montenegrin-language version), Shanghai-China, Serbia (Serbian-language version), Chinese Taipei |
| R104Q01 | Ireland (booklet 11 of English-language version) |
| R104Q02 | Lithuania (Lithuanian-language version), Montenegro (Serbian/variant of Montenegrin, All) |
| R111Q02B | Qatar (booklet 8 of Arabic-language version) |
| R111Q06B | Peru (booklet 24) |
| R219Q01E | International Deletion |
| R219Q01T | International Deletion |
| R220Q02B | Brazil, Switzerland (booklet 9 Italian-language version), Indonesia (booklet 9), Japan, Peru |
| R220Q04 | Hungary, Indonesia (booklets 1 and 9) |
| R220Q05 | Bulgaria (booklet 13), Spain (Catalan-language version), Portugal (booklets 2 and 13 of Portuguese-language version) |
| R220Q06 | Estonia (Russian-language version) |
| R227Q01 | Montenegro (Serbian/variant of Montenegrin-language version) |
| R227Q02 | Azerbaijan (booklet 11 of Azerbaijani), Finland, Israel (Arabic-language version), Montenegro (Serbian/variant of Montenegrin-language version) |
| R227Q03 | Israel (Arabic-language version), Montenegro (Serbian/variant of Montenegrin-language version) |
| R227Q06 | Kazakhstan (Russian-language version), Dubai (UAE) (Arabic-language version) |
| R245Q01 | Greece, Israel (booklet 5 of Arabic-language version), Slovak Republic (Hungarian-language version) |
| R245Q02 | Iceland, Israel (booklet 5 of Arabic-language version) |
| R412Q05 | Dubai (UAE) (Arabic-language version) |
| R412Q06T | Chile |
| R414Q02 | Poland |
| R414Q09 | Switzerland (booklets 4 and 6 of French-language version) |
| R420Q09 | Estonia (Estonian-language version) |
| R420Q10 | Japan (booklet 6) |
| R424Q02T | Argentina, Montenegro (Serbian/variant of Montenegrin-language version) |
| R432Q05 | Turkey (booklet 2) |
| R432Q06T | Ireland, Kazakhstan (booklet 12 of Russian-language version), Lithuania, Singapore |
| R433Q02 | Chile |



[Part 2/2]

Table 12.8 Items deleted at the national level

| Item | Country |
|----------|---|
| R437Q01 | Brazil |
| R437Q06 | Hungary |
| R437Q07 | Chile |
| R442Q06 | Israel (booklet 7 of Arabic-language version) |
| R442Q07 | Hungary (booklet 5) |
| R445Q03 | Romania (all booklets of Hungarian-language version) |
| R452Q06 | Austria |
| R453Q01 | Argentina, Qatar |
| R453Q04 | Argentina |
| R453Q05T | Argentina |
| R453Q06 | Argentina, Iceland |
| R455Q03 | Greece |
| R455Q05T | Austria (German-language version), Belgium (German-language version), Switzerland (German-language version), Germany, Italy (German-language version), Luxembourg (German-language version) |
| R462Q02 | Serbia (booklets 22 and 24 of Serbian-language version) |
| R462Q04 | Serbia (booklet 24 of Serbian-language version) |
| R462Q05 | Serbia (booklet 24 of Serbian-language version) |
| R466Q03T | Albania, Poland, Serbia (booklet 26 of Serbian-language version), Trinidad and Tobago (booklet 22), Tunisia |
| R466Q06 | Argentina (booklets 11 and 12), Qatar (booklet 26 of English-language version), Serbia (Hungarian-language version), Trinidad and Tobago (booklet 22) |
| S326Q03 | Croatia |
| S413Q04T | Colombia |
| S425Q02 | Tunisia (booklet 23 of Arabic-language version) |
| S425Q05 | Croatia |
| S438Q03D | Israel (Arabic-language version) |
| S465Q01 | Spain (Euskara-language version) |
| S466Q05 | Peru (booklet 24) |
| S478Q01 | Dubai (UAE) (Arabic-language version) |
| S478Q02T | Dubai (UAE) (Arabic-language version), Uruguay (booklet 12) |
| S498Q04 | Peru (booklet 13) |
| S519Q01 | Peru |
| S519Q03 | Israel (Hebrew-language version) |
| S527Q04T | Macao-China (Cantonese-language version) |
| E002Q01 | Sweden |
| E017Q01 | Norway |
| E017Q07 | Iceland |
| E021Q05 | Sweden |
| E021Q08 | Iceland |

International scaling

The international scaling for mathematics, science and paper-based reading items were performed using a calibration data set of 15 500 students (500 randomly selected students from each of the 31 OECD countries). For the estimation of non-standard reading international item parameters a calibration sample of 24 500 students was used. This calibration sample included 500 students from all OECD countries and 500 students from 20 countries that administered the non-standard test.

The item parameter estimates from this scaling are reported in Annex A. The item parameters were estimated using three separate one-dimensional models. As in previous cycles, not-reached items were treated as not administered and a booklet facet was used in the item response model.

The international scaling for DRA items was performed using calibration data set of 4 370 students (230 randomly selected students from each of the 19 participating countries). The item parameter estimates from this scaling are reported in Annex A.

Generating student scale scores and reliability of the PISA scales

Applying the conditioning approach described in Chapter 9 and anchoring all of the item parameters at the values obtained from the international scaling, plausible values were generated for all sampled students. Table 12.9 gives the reliabilities at the international level for the generated scale scores. The increase in reliability of the results reported in Table 12.9 over those presented in Table 12.3 is due to the use of multidimensional scaling and conditioning.

Table 12.10 gives the reliabilities at the national level for the generated scale scores.

Table 12.9 Final reliabilities of the PISA scales

| Domain | Reliability |
|-------------------------|-------------|
| Mathematics | 0.882 |
| Reading | 0.921 |
| Science | 0.896 |
| Access and retrieve | 0.907 |
| Integrate and interpret | 0.913 |
| Reflect and evaluate | 0.909 |
| Continuous text | 0.911 |
| Non-continuous text | 0.903 |
| DRA | 0.900 |

[Part 1/2]

Table 12.10 National reliabilities of the PISA scales

| | Mathematics | Reading | Science | Access and retrieve | Integrate and interpret | Reflect and evaluate | Continuous text | Non-continuous text | DRA | |
|---------------|-----------------|---------|---------|---------------------|-------------------------|----------------------|-----------------|---------------------|------|------|
| OECD | Australia | 0.88 | 0.93 | 0.91 | 0.91 | 0.93 | 0.92 | 0.92 | 0.90 | 0.91 |
| | Austria | 0.89 | 0.93 | 0.92 | 0.91 | 0.92 | 0.93 | 0.92 | 0.91 | 0.93 |
| | Belgium | 0.92 | 0.94 | 0.93 | 0.92 | 0.93 | 0.92 | 0.93 | 0.93 | 0.93 |
| | Canada | 0.87 | 0.91 | 0.89 | 0.90 | 0.91 | 0.90 | 0.90 | 0.89 | 0.89 |
| | Chile | 0.87 | 0.90 | 0.86 | 0.87 | 0.89 | 0.88 | 0.88 | 0.88 | |
| | Czech Republic | 0.89 | 0.92 | 0.89 | 0.91 | 0.92 | 0.91 | 0.91 | 0.92 | |
| | Denmark | 0.86 | 0.91 | 0.90 | 0.90 | 0.91 | 0.91 | 0.90 | 0.88 | 0.93 |
| | Estonia | 0.86 | 0.91 | 0.87 | 0.89 | 0.90 | 0.90 | 0.90 | 0.89 | |
| | Finland | 0.83 | 0.90 | 0.87 | 0.88 | 0.89 | 0.89 | 0.89 | 0.88 | |
| | France | 0.89 | 0.94 | 0.91 | 0.92 | 0.93 | 0.92 | 0.93 | 0.91 | 0.94 |
| | Germany | 0.91 | 0.92 | 0.92 | 0.92 | 0.92 | 0.93 | 0.91 | 0.91 | |
| | Greece | 0.84 | 0.91 | 0.86 | 0.89 | 0.89 | 0.89 | 0.90 | 0.89 | |
| | Hungary | 0.91 | 0.93 | 0.92 | 0.91 | 0.92 | 0.91 | 0.92 | 0.92 | 0.92 |
| | Iceland | 0.87 | 0.92 | 0.90 | 0.90 | 0.91 | 0.91 | 0.91 | 0.90 | 0.90 |
| | Ireland | 0.90 | 0.93 | 0.91 | 0.93 | 0.93 | 0.92 | 0.92 | 0.91 | 0.89 |
| | Israel | 0.89 | 0.93 | 0.89 | 0.94 | 0.92 | 0.93 | 0.92 | 0.91 | |
| | Italy | 0.89 | 0.93 | 0.90 | 0.91 | 0.92 | 0.92 | 0.92 | 0.91 | |
| | Japan | 0.89 | 0.92 | 0.91 | 0.91 | 0.91 | 0.89 | 0.91 | 0.90 | 0.84 |
| | Korea | 0.87 | 0.91 | 0.90 | 0.91 | 0.90 | 0.89 | 0.89 | 0.90 | 0.87 |
| | Luxembourg | 0.87 | 0.93 | 0.90 | 0.90 | 0.92 | 0.91 | 0.92 | 0.92 | |
| | Mexico | 0.88 | 0.91 | 0.87 | 0.90 | 0.90 | 0.90 | 0.90 | 0.90 | |
| | Netherlands | 0.91 | 0.93 | 0.92 | 0.92 | 0.93 | 0.93 | 0.92 | 0.92 | |
| | New Zealand | 0.89 | 0.94 | 0.92 | 0.92 | 0.93 | 0.92 | 0.93 | 0.92 | 0.90 |
| | Norway | 0.86 | 0.92 | 0.88 | 0.90 | 0.90 | 0.90 | 0.91 | 0.89 | 0.86 |
| | Poland | 0.88 | 0.92 | 0.88 | 0.89 | 0.91 | 0.90 | 0.90 | 0.89 | 0.90 |
| | Portugal | 0.88 | 0.91 | 0.87 | 0.89 | 0.90 | 0.90 | 0.90 | 0.88 | |
| | Slovak Republic | 0.89 | 0.93 | 0.89 | 0.92 | 0.91 | 0.92 | 0.91 | 0.92 | |
| | Slovenia | 0.89 | 0.93 | 0.90 | 0.93 | 0.93 | 0.92 | 0.92 | 0.92 | |
| | Spain | 0.89 | 0.92 | 0.90 | 0.90 | 0.92 | 0.91 | 0.91 | 0.91 | 0.89 |
| | Sweden | 0.87 | 0.92 | 0.89 | 0.91 | 0.91 | 0.92 | 0.91 | 0.90 | 0.90 |
| | Switzerland | 0.88 | 0.92 | 0.89 | 0.90 | 0.91 | 0.91 | 0.91 | 0.90 | |
| | Turkey | 0.90 | 0.91 | 0.87 | 0.90 | 0.90 | 0.88 | 0.90 | 0.89 | |
| | United Kingdom | 0.88 | 0.92 | 0.91 | 0.91 | 0.92 | 0.90 | 0.92 | 0.89 | |
| United States | 0.88 | 0.92 | 0.91 | 0.91 | 0.92 | 0.91 | 0.91 | 0.91 | | |



[Part 2/2]
Table 12.10 National reliabilities of the PISA scales

| | Mathematics | Reading | Science | Access and retrieve | Integrate and interpret | Reflect and evaluate | Continuous text | Non-continuous text | DRA | |
|-----------------|---------------------|---------|---------|---------------------|-------------------------|----------------------|-----------------|---------------------|------|------|
| <i>Partners</i> | Albania | 0.84 | 0.92 | 0.87 | 0.91 | 0.90 | 0.91 | 0.90 | | |
| | Argentina | 0.88 | 0.92 | 0.89 | 0.91 | 0.92 | 0.91 | 0.91 | | |
| | Azerbaijan | 0.77 | 0.86 | 0.79 | 0.87 | 0.84 | 0.84 | 0.84 | | |
| | Brazil | 0.90 | 0.91 | 0.89 | 0.91 | 0.91 | 0.91 | 0.91 | 0.92 | |
| | Bulgaria | 0.88 | 0.94 | 0.89 | 0.92 | 0.93 | 0.91 | 0.93 | 0.93 | |
| | Colombia | 0.86 | 0.90 | 0.85 | 0.89 | 0.89 | 0.89 | 0.89 | 0.90 | 0.90 |
| | Croatia | 0.89 | 0.92 | 0.89 | 0.91 | 0.92 | 0.91 | 0.91 | 0.91 | |
| | Dubai (UAE) | 0.88 | 0.93 | 0.90 | 0.93 | 0.92 | 0.92 | 0.92 | 0.92 | |
| | Hong Kong-China | 0.88 | 0.91 | 0.89 | 0.91 | 0.90 | 0.90 | 0.90 | 0.90 | 0.87 |
| | Indonesia | 0.80 | 0.86 | 0.81 | 0.88 | 0.84 | 0.81 | 0.85 | 0.82 | |
| | Jordan | 0.88 | 0.91 | 0.86 | 0.91 | 0.91 | 0.90 | 0.88 | 0.87 | |
| | Kazakhstan | 0.85 | 0.91 | 0.85 | 0.90 | 0.90 | 0.90 | 0.87 | 0.87 | |
| | Kyrgyzstan | 0.82 | 0.90 | 0.82 | 0.89 | 0.88 | 0.88 | 0.89 | 0.88 | |
| | Latvia | 0.86 | 0.91 | 0.87 | 0.91 | 0.89 | 0.89 | 0.89 | 0.91 | |
| | Liechtenstein | 0.91 | 0.93 | 0.92 | 0.93 | 0.92 | 0.93 | 0.92 | 0.92 | |
| | Lithuania | 0.89 | 0.92 | 0.89 | 0.92 | 0.91 | 0.92 | 0.91 | 0.90 | |
| | Macao-China | 0.82 | 0.89 | 0.83 | 0.87 | 0.87 | 0.86 | 0.87 | 0.84 | 0.79 |
| | Montenegro | 0.85 | 0.91 | 0.84 | 0.89 | 0.90 | 0.88 | 0.89 | 0.88 | |
| | Panama | 0.89 | 0.92 | 0.87 | 0.91 | 0.91 | 0.90 | 0.91 | 0.91 | |
| | Peru | 0.86 | 0.91 | 0.83 | 0.89 | 0.90 | 0.90 | 0.90 | 0.90 | |
| | Qatar | 0.87 | 0.93 | 0.89 | 0.93 | 0.92 | 0.91 | 0.92 | 0.92 | |
| | Romania | 0.86 | 0.92 | 0.87 | 0.92 | 0.91 | 0.91 | 0.90 | 0.91 | |
| | Russian Federation | 0.86 | 0.91 | 0.85 | 0.90 | 0.90 | 0.90 | 0.90 | 0.89 | |
| | Serbia | 0.86 | 0.91 | 0.87 | 0.90 | 0.90 | 0.89 | 0.89 | 0.90 | |
| | Shanghai-China | 0.86 | 0.89 | 0.86 | 0.88 | 0.89 | 0.89 | 0.89 | 0.87 | |
| | Singapore | 0.89 | 0.93 | 0.91 | 0.89 | 0.92 | 0.91 | 0.93 | 0.91 | |
| | Chinese Taipei | 0.88 | 0.91 | 0.89 | 0.90 | 0.91 | 0.90 | 0.91 | 0.89 | |
| | Thailand | 0.84 | 0.89 | 0.84 | 0.87 | 0.88 | 0.88 | 0.89 | 0.88 | |
| | Trinidad and Tobago | 0.89 | 0.93 | 0.89 | 0.92 | 0.92 | 0.93 | 0.92 | 0.92 | |
| | Tunisia | 0.82 | 0.89 | 0.83 | 0.89 | 0.87 | 0.88 | 0.88 | 0.87 | |
| Uruguay | 0.86 | 0.91 | 0.85 | 0.89 | 0.91 | 0.89 | 0.90 | 0.89 | | |

TEST LENGTH ANALYSIS

Numbers of missing and non reached responses are discussed in this section. A response is coded as missing if the student was expected to answer a question, but no response was actually provided. All consecutive missing values clustered at the end of a test session were replaced by the non-reached code, except for the first value of the missing series, which is coded as missing (see Chapter 18). All the tables included in the section include weighted and unweighted numbers of the missing and not-reached responses. Final student weight (see Chapter 8) was used to provide weighted numbers and percents.

Table 12.11 shows the number of missing responses and the number of missing responses recoded as not reached, by booklet. Table 12.12 shows the number of missing and not-reached responses by DRA test form.

Table 12.11 Average number of not-reached items and missing items by booklet

| Booklet | Missing | | Not reached | |
|--------------|-------------|-------------|-------------|-------------|
| | Weighted | Unweighted | Weighted | Unweighted |
| 1 | 5.05 | 5.26 | 0.76 | 0.69 |
| 2 | 3.91 | 3.98 | 1.39 | 1.09 |
| 3 | 4.56 | 4.98 | 1.26 | 1.10 |
| 4 | 4.27 | 4.65 | 1.81 | 1.36 |
| 5 | 4.57 | 4.79 | 0.82 | 0.64 |
| 6 | 3.38 | 3.55 | 1.21 | 0.85 |
| 7 | 4.48 | 4.72 | 1.38 | 1.24 |
| 8 | 5.61 | 6.17 | 2.73 | 2.54 |
| 9 | 4.23 | 4.79 | 1.45 | 1.53 |
| 10 | 5.02 | 5.56 | 2.21 | 2.02 |
| 11 | 4.84 | 5.38 | 1.47 | 1.36 |
| 12 | 4.50 | 4.88 | 2.34 | 2.20 |
| 13 | 4.82 | 5.33 | 2.64 | 2.55 |
| 21 | 5.97 | 6.52 | 2.01 | 1.89 |
| 22 | 4.42 | 5.12 | 3.51 | 3.20 |
| 23 | 5.52 | 6.07 | 3.30 | 3.09 |
| 24 | 3.39 | 3.93 | 2.38 | 2.34 |
| 25 | 5.50 | 5.89 | 1.73 | 1.61 |
| 26 | 4.08 | 4.82 | 4.07 | 3.46 |
| 27 | 5.35 | 6.00 | 4.22 | 3.76 |
| UH | 3.99 | 3.78 | 0.93 | 1.20 |
| Total | 4.64 | 5.09 | 1.91 | 1.79 |

Average number of missing and not-reached items could be compared between standard booklets 1 to 7 and non-standard booklets 21 to 27. Standard booklets have on average less not-reached items and less missing data.

Table 12.12 Average number of not-reached items and missing items by DRA TestID

| TestID | Missing | | Not reached | |
|--------------|-------------|-------------|-------------|-------------|
| | Weighted | Unweighted | Weighted | Unweighted |
| 1 | 1.64 | 1.52 | 0.48 | 0.41 |
| 2 | 1.87 | 1.72 | 0.45 | 0.41 |
| 3 | 1.55 | 1.39 | 0.36 | 0.28 |
| 4 | 1.36 | 1.18 | 0.36 | 0.32 |
| 5 | 1.17 | 0.96 | 0.37 | 0.30 |
| 6 | 1.22 | 1.09 | 0.30 | 0.24 |
| Total | 1.47 | 1.31 | 0.39 | 0.33 |

Table 12.13 shows the number of not-reached items for the paper and pencil assessment, by country. Table 12.14 shows this information by country over all booklets and DRA test form. The average number of not-reached items differs from one country to another. Generally, countries with higher averages of not-reached items also have higher averages of missing data. Tables 12.15 and 12.16 provide the percentage distribution of not-reached items per booklet and DRA test form. The percentage of students who reached the last item (i.e. the percentages of students with zero not-reached items) for paper and pencil assessment ranges from 67% to 91% when using weighted data and 68% to 91% when using unweighted data. The percentage of students who reached the last item for DRA assessment ranges from 89% to 91% when using weighted data and 90% to 93% when using unweighted data.

Table 12.13 Average number of not-reached items and missing items by country

| | Missing | | Not reached | |
|---------------------|----------|------------|-------------|------------|
| | Weighted | Unweighted | Weighted | Unweighted |
| OECD | | | | |
| Australia | 3.23 | 3.58 | 0.90 | 1.11 |
| Austria | 6.69 | 6.43 | 0.54 | 0.56 |
| Belgium | 4.08 | 4.03 | 0.93 | 0.90 |
| Canada | 2.68 | 2.99 | 0.85 | 0.85 |
| Chile | 5.44 | 5.28 | 2.08 | 2.07 |
| Czech Republic | 6.20 | 5.34 | 0.70 | 0.60 |
| Denmark | 4.71 | 5.35 | 0.88 | 1.03 |
| Estonia | 4.08 | 4.05 | 0.61 | 0.61 |
| Finland | 2.83 | 2.88 | 0.43 | 0.49 |
| France | 6.58 | 6.48 | 1.77 | 1.72 |
| Germany | 5.36 | 5.44 | 0.67 | 0.69 |
| Greece | 6.42 | 6.14 | 1.56 | 1.56 |
| Hungary | 4.95 | 4.70 | 0.50 | 0.45 |
| Iceland | 3.80 | 3.81 | 1.19 | 1.18 |
| Ireland | 4.09 | 4.01 | 1.07 | 1.04 |
| Israel | 6.84 | 6.78 | 2.33 | 2.24 |
| Italy | 5.84 | 5.58 | 1.53 | 1.28 |
| Japan | 5.01 | 4.94 | 0.71 | 0.68 |
| Korea | 2.42 | 2.32 | 0.19 | 0.16 |
| Luxembourg | 6.55 | 6.33 | 1.41 | 1.31 |
| Mexico | 3.00 | 2.92 | 3.46 | 3.38 |
| Netherlands | 1.44 | 1.31 | 0.19 | 0.18 |
| New Zealand | 3.27 | 3.21 | 1.03 | 1.01 |
| Norway | 4.79 | 4.79 | 1.09 | 1.10 |
| Poland | 4.35 | 4.15 | 0.54 | 0.55 |
| Portugal | 4.52 | 4.63 | 1.39 | 1.41 |
| Slovak Republic | 5.66 | 5.62 | 0.62 | 0.61 |
| Slovenia | 5.49 | 6.62 | 0.32 | 0.48 |
| Spain | 5.22 | 5.00 | 1.58 | 1.45 |
| Sweden | 5.04 | 5.00 | 1.51 | 1.49 |
| Switzerland | 4.66 | 4.68 | 0.59 | 0.66 |
| Turkey | 4.61 | 4.57 | 0.95 | 0.89 |
| United Kingdom | 3.98 | 4.24 | 0.70 | 0.63 |
| United States | 1.68 | 1.71 | 0.56 | 0.57 |
| Partners | | | | |
| Albania | 12.58 | 12.68 | 2.79 | 2.68 |
| Argentina | 9.33 | 9.29 | 5.87 | 5.50 |
| Azerbaijan | 13.66 | 13.41 | 1.43 | 1.35 |
| Brazil | 4.69 | 4.92 | 2.66 | 3.01 |
| Bulgaria | 8.97 | 8.96 | 1.92 | 2.11 |
| Colombia | 4.58 | 4.42 | 6.21 | 5.36 |
| Croatia | 5.52 | 5.56 | 0.43 | 0.43 |
| Dubai (UAE) | 4.16 | 4.60 | 1.29 | 1.40 |
| Hong Kong-China | 2.48 | 2.44 | 0.40 | 0.39 |
| Indonesia | 6.44 | 6.42 | 2.98 | 3.00 |
| Jordan | 5.46 | 5.02 | 2.13 | 1.98 |
| Kazakhstan | 7.40 | 7.23 | 3.77 | 3.61 |
| Kyrgyzstan | 12.44 | 12.23 | 8.89 | 8.74 |
| Latvia | 3.70 | 3.55 | 0.93 | 0.87 |
| Liechtenstein | 4.32 | 4.37 | 0.83 | 0.82 |
| Lithuania | 4.73 | 4.71 | 0.63 | 0.61 |
| Macao-China | 3.30 | 3.29 | 1.20 | 1.20 |
| Montenegro | 11.48 | 11.63 | 1.50 | 1.48 |
| Panama | 7.36 | 7.03 | 4.18 | 4.50 |
| Peru | 7.84 | 7.67 | 6.28 | 6.26 |
| Qatar | 7.57 | 7.52 | 2.00 | 1.95 |
| Romania | 4.36 | 4.44 | 0.76 | 0.78 |
| Russian Federation | 6.09 | 6.11 | 2.62 | 2.56 |
| Serbia | 7.96 | 7.94 | 1.00 | 1.05 |
| Shanghai-China | 1.29 | 1.30 | 0.10 | 0.11 |
| Singapore | 2.49 | 2.55 | 0.67 | 0.69 |
| Chinese Taipei | 3.34 | 3.27 | 0.45 | 0.46 |
| Thailand | 3.86 | 3.75 | 1.22 | 1.19 |
| Trinidad and Tobago | 7.47 | 7.53 | 4.90 | 4.73 |
| Tunisia | 7.61 | 7.86 | 3.48 | 3.51 |
| Uruguay | 8.09 | 8.17 | 4.56 | 4.63 |

Table 12.14 Average number of DRA not-reached items and missing items by country

| | Missing | | Not reached | |
|-----------------|----------|------------|-------------|------------|
| | Weighted | Unweighted | Weighted | Unweighted |
| OECD | | | | |
| Australia | 0.77 | 0.86 | 0.11 | 0.14 |
| Austria | 1.87 | 1.89 | 0.16 | 0.23 |
| Belgium | 1.11 | 1.08 | 0.13 | 0.14 |
| Chile | 2.38 | 2.29 | 1.19 | 1.16 |
| Denmark | 0.97 | 1.19 | 0.13 | 0.15 |
| France | 1.29 | 1.24 | 0.43 | 0.43 |
| Hungary | 1.89 | 1.65 | 0.16 | 0.15 |
| Iceland | 0.85 | 0.84 | 0.14 | 0.14 |
| Ireland | 1.08 | 1.06 | 0.19 | 0.17 |
| Japan | 1.32 | 1.31 | 0.27 | 0.28 |
| Korea | 0.53 | 0.50 | 0.08 | 0.08 |
| New Zealand | 0.85 | 0.83 | 0.22 | 0.23 |
| Norway | 1.01 | 1.02 | 0.19 | 0.18 |
| Poland | 2.05 | 1.93 | 0.11 | 0.11 |
| Spain | 1.47 | 1.45 | 0.18 | 0.17 |
| Sweden | 1.07 | 1.07 | 0.19 | 0.19 |
| Partners | | | | |
| Colombia | 2.87 | 2.74 | 1.62 | 1.64 |
| Hong Kong-China | 1.00 | 1.01 | 0.34 | 0.35 |
| Macao-China | 1.01 | 1.01 | 0.60 | 0.60 |

Table 12.15 Distribution of not-reached items by booklet

| Booklet | Number of non-reached items | | | | | | | | | |
|---------|-----------------------------|------|------|------|------|------|------|------|------|-------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | >8 |
| | Weighted percentages | | | | | | | | | |
| 1 | 91.20 | 0.28 | 1.03 | 1.54 | 0.21 | 0.73 | 0.90 | 0.67 | 0.13 | 3.44 |
| 2 | 86.86 | 0.85 | 1.79 | 1.07 | 0.83 | 0.84 | 0.33 | 0.48 | 1.09 | 6.94 |
| 3 | 86.94 | 1.08 | 1.42 | 1.64 | 1.01 | 0.60 | 0.65 | 0.71 | 0.17 | 5.95 |
| 4 | 79.48 | 1.28 | 3.62 | 1.19 | 0.84 | 3.69 | 0.81 | 0.92 | 0.59 | 8.18 |
| 5 | 86.87 | 2.06 | 2.83 | 2.21 | 0.64 | 1.41 | 0.10 | 0.27 | 0.15 | 3.60 |
| 6 | 88.76 | 0.81 | 0.73 | 1.06 | 0.23 | 0.67 | 0.38 | 1.87 | 0.30 | 5.50 |
| 7 | 87.90 | 0.27 | 0.82 | 0.21 | 1.48 | 0.38 | 1.35 | 0.14 | 0.42 | 7.44 |
| 8 | 78.70 | 1.22 | 3.09 | 0.33 | 1.11 | 0.63 | 0.94 | 1.36 | 0.40 | 12.62 |
| 9 | 85.40 | 0.42 | 0.57 | 1.04 | 2.40 | 0.40 | 0.79 | 1.41 | 0.89 | 7.58 |
| 10 | 81.83 | 1.29 | 0.68 | 0.94 | 1.91 | 0.71 | 0.91 | 0.76 | 0.75 | 10.97 |
| 11 | 87.26 | 0.57 | 0.52 | 0.55 | 1.28 | 0.30 | 0.67 | 1.03 | 0.56 | 7.82 |
| 12 | 82.72 | 0.76 | 1.27 | 0.58 | 0.68 | 0.61 | 1.01 | 0.46 | 1.49 | 11.90 |
| 13 | 80.58 | 0.64 | 0.63 | 0.88 | 1.03 | 0.90 | 1.13 | 0.29 | 1.53 | 13.92 |
| 21 | 79.82 | 1.21 | 1.18 | 2.77 | 0.65 | 0.95 | 2.32 | 1.94 | 0.47 | 9.15 |
| 22 | 72.58 | 1.52 | 2.71 | 2.11 | 1.10 | 1.78 | 0.67 | 0.82 | 1.92 | 16.71 |
| 23 | 72.43 | 1.11 | 1.90 | 4.03 | 1.73 | 1.29 | 1.09 | 1.20 | 0.86 | 15.21 |
| 24 | 72.43 | 1.83 | 3.75 | 1.72 | 1.34 | 4.51 | 1.51 | 1.19 | 1.36 | 11.74 |
| 25 | 76.52 | 2.58 | 4.01 | 2.67 | 1.27 | 2.33 | 0.21 | 0.95 | 1.03 | 9.46 |
| 26 | 74.86 | 0.87 | 0.88 | 0.90 | 0.39 | 0.48 | 0.63 | 1.42 | 0.36 | 19.58 |
| 27 | 67.64 | 2.92 | 1.99 | 3.40 | 1.35 | 1.67 | 0.42 | 1.76 | 0.55 | 18.86 |
| UH | 85.42 | 2.80 | 0.31 | 1.74 | 0.68 | 2.67 | 0.33 | 0.50 | 1.38 | 5.55 |
| | Unweighted percentages | | | | | | | | | |
| 1 | 91.05 | 0.41 | 1.24 | 1.71 | 0.39 | 0.66 | 1.11 | 0.73 | 0.22 | 2.49 |
| 2 | 88.79 | 0.79 | 1.98 | 1.10 | 0.69 | 0.94 | 0.25 | 0.36 | 0.68 | 4.42 |
| 3 | 87.53 | 1.06 | 1.58 | 2.09 | 0.82 | 0.54 | 0.64 | 0.62 | 0.27 | 4.85 |
| 4 | 82.26 | 1.71 | 3.12 | 1.43 | 0.76 | 3.39 | 0.76 | 0.69 | 0.55 | 5.33 |
| 5 | 88.50 | 2.38 | 2.35 | 1.93 | 0.67 | 1.01 | 0.13 | 0.32 | 0.18 | 2.53 |
| 6 | 91.01 | 0.95 | 0.72 | 1.01 | 0.18 | 0.58 | 0.30 | 1.52 | 0.16 | 3.57 |
| 7 | 88.61 | 0.35 | 0.80 | 0.20 | 1.64 | 0.53 | 1.36 | 0.14 | 0.36 | 6.01 |
| 8 | 79.62 | 1.44 | 2.89 | 0.32 | 1.33 | 0.60 | 0.97 | 1.26 | 0.32 | 11.24 |
| 9 | 84.38 | 0.53 | 0.35 | 0.87 | 3.54 | 0.19 | 0.60 | 1.66 | 0.89 | 7.00 |
| 10 | 82.84 | 1.31 | 0.56 | 0.76 | 1.99 | 0.61 | 0.86 | 0.70 | 0.75 | 9.62 |
| 11 | 88.15 | 0.44 | 0.57 | 0.46 | 1.30 | 0.37 | 0.54 | 1.13 | 0.56 | 6.49 |
| 12 | 83.45 | 0.77 | 1.16 | 0.58 | 0.64 | 0.84 | 1.04 | 0.49 | 1.38 | 9.64 |
| 13 | 81.15 | 0.63 | 0.63 | 0.98 | 0.99 | 0.71 | 1.24 | 0.32 | 1.90 | 11.45 |
| 21 | 80.94 | 0.97 | 1.38 | 2.94 | 0.53 | 1.03 | 2.19 | 1.57 | 0.52 | 7.94 |
| 22 | 74.32 | 1.54 | 2.79 | 2.19 | 1.22 | 1.52 | 0.68 | 0.73 | 1.71 | 13.29 |
| 23 | 73.55 | 1.20 | 2.07 | 3.50 | 1.57 | 1.29 | 1.14 | 1.22 | 0.63 | 13.83 |
| 24 | 72.67 | 1.53 | 3.99 | 1.76 | 1.00 | 4.89 | 1.40 | 1.33 | 1.18 | 10.25 |
| 25 | 77.25 | 2.99 | 4.20 | 2.90 | 1.15 | 1.97 | 0.18 | 0.83 | 0.80 | 7.73 |
| 26 | 77.15 | 1.05 | 0.82 | 0.95 | 0.36 | 0.41 | 0.77 | 1.54 | 0.45 | 16.49 |
| 27 | 68.22 | 3.17 | 2.46 | 4.65 | 1.51 | 1.39 | 0.31 | 1.82 | 0.43 | 16.04 |
| UH | 83.39 | 2.02 | 0.83 | 2.20 | 1.47 | 2.57 | 0.46 | 0.28 | 0.83 | 5.96 |



Table 12.16 Distribution of not-reached items by DRA TestID

| Booklet | Number of non-reached items | | | | | | | | | |
|---------|-----------------------------|------|------|------|------|------|------|------|------|------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | >8 |
| | Weighted percentages | | | | | | | | | |
| 1 | 89.08 | 1.95 | 1.14 | 1.66 | 2.29 | 1.13 | 0.86 | 0.44 | 0.47 | 0.97 |
| 2 | 89.29 | 2.62 | 2.11 | 1.08 | 1.10 | 0.73 | 1.01 | 0.56 | 0.07 | 1.43 |
| 3 | 90.74 | 1.89 | 2.36 | 0.92 | 1.29 | 0.96 | 0.62 | 0.17 | 0.08 | 0.95 |
| 4 | 89.97 | 2.83 | 1.14 | 1.09 | 2.17 | 0.97 | 0.66 | 0.57 | 0.28 | 0.32 |
| 5 | 90.31 | 2.17 | 2.23 | 1.18 | 0.84 | 0.96 | 1.10 | 0.25 | 0.17 | 0.78 |
| 6 | 91.33 | 1.96 | 2.53 | 1.45 | 0.79 | 0.68 | 0.20 | 0.17 | 0.24 | 0.64 |
| | Unweighted percentages | | | | | | | | | |
| 1 | 90.14 | 1.87 | 1.02 | 1.53 | 2.13 | 0.85 | 0.98 | 0.46 | 0.30 | 0.72 |
| 2 | 90.12 | 2.22 | 2.04 | 1.21 | 0.91 | 0.78 | 1.06 | 0.55 | 0.11 | 0.99 |
| 3 | 92.01 | 2.03 | 2.17 | 0.75 | 1.00 | 0.80 | 0.42 | 0.12 | 0.10 | 0.60 |
| 4 | 91.45 | 2.18 | 1.05 | 0.91 | 1.93 | 0.81 | 0.60 | 0.38 | 0.28 | 0.40 |
| 5 | 91.63 | 2.10 | 1.85 | 1.07 | 0.91 | 0.81 | 0.79 | 0.21 | 0.10 | 0.53 |
| 6 | 92.74 | 1.76 | 1.92 | 1.10 | 0.82 | 0.79 | 0.18 | 0.11 | 0.18 | 0.39 |

BOOKLET EFFECTS

The booklet parameters for the paper and pencil test that are described in Chapter 9 are reported in Table 12.17. The booklet effects are the amount that must be added to the proficiencies of students who responded to each booklet. That is, a positive value indicates a booklet that was harder than the average while a negative value indicates a booklet that was easier than the average. Since the booklet effects are deviations from an average they sum to zero for each domain. Table 12.18 shows the booklet effects after transformation to the PISA scales.

Table 12.17 Estimated booklet effects in logits

| Booklet | Domains | | |
|--------------|-------------|---------|---------|
| | Mathematics | Reading | Science |
| Standard set | | | |
| 1 | 0.045 | -0.04 | |
| 2 | | 0.068 | -0.1 |
| 3 | 0.069 | -0.045 | 0.008 |
| 4 | | 0.027 | 0.097 |
| 5 | 0.064 | -0.181 | |
| 6 | | -0.056 | |
| 7 | -0.058 | 0.101 | -0.027 |
| 8 | -0.071 | 0.122 | 0.071 |
| 9 | -0.13 | 0.251 | -0.032 |
| 10 | 0.045 | -0.091 | 0.107 |
| 11 | 0.024 | -0.026 | |
| 12 | 0.011 | -0.387 | 0.068 |
| 13 | | 0.258 | -0.193 |
| Easy set | | | |
| 8 | 0.037 | 0.276 | 0.38 |
| 9 | -0.248 | 0.387 | -0.18 |
| 10 | 0.033 | -0.238 | 0.228 |
| 11 | 0.092 | 0.132 | |
| 12 | 0.204 | -0.289 | 0.105 |
| 13 | | 0.406 | -0.336 |
| 21 | 0.022 | -0.173 | |
| 22 | | 0.127 | -0.026 |
| 23 | 0.006 | -0.299 | -0.016 |
| 24 | | -0.196 | -0.12 |
| 25 | 0.014 | -0.348 | |
| 26 | | 0.060 | |
| 27 | -0.16 | 0.156 | -0.035 |

Table 12.18 Estimated booklet effects on the PISA scale

| Booklet | Domains | | |
|---------------------|-------------|---------|---------|
| | Mathematics | Reading | Science |
| Standard set | | | |
| 1 | 3.5 | -3.2 | |
| 2 | | 5.5 | -9.3 |
| 3 | 5.4 | -3.6 | 0.7 |
| 4 | | 2.2 | 9.0 |
| 5 | 5.0 | -14.5 | |
| 6 | | -4.5 | |
| 7 | -4.5 | 8.1 | -2.5 |
| 8 | -5.5 | 9.8 | 6.6 |
| 9 | -10.1 | 20.1 | -3.0 |
| 10 | 3.5 | -7.3 | 10.0 |
| 11 | 1.9 | -2.1 | |
| 12 | 0.9 | -31.1 | 6.3 |
| 13 | | 20.7 | -18.0 |
| Easy set | | | |
| 8 | 2.9 | 22.1 | 35.4 |
| 9 | -19.3 | 31.1 | -16.8 |
| 10 | 2.6 | -19.1 | 21.3 |
| 11 | 7.2 | 10.6 | |
| 12 | 15.9 | -23.2 | 9.8 |
| 13 | | 32.6 | -31.3 |
| 21 | 1.7 | -13.9 | |
| 22 | | 10.2 | -2.4 |
| 23 | 0.5 | -24.0 | -1.5 |
| 24 | | -15.7 | -11.2 |
| 25 | 1.1 | -27.9 | |
| 26 | | 4.8 | |
| 27 | -12.5 | 12.5 | -3.3 |

Booklets that include a single domain cluster at the beginning of the booklet (mathematics in booklet 9, reading in booklet 12 and science in booklet 13) have the largest negative parameters. Booklets with the domain at the end of the booklet have the highest positive parameters. The reading booklet effects for the non-standard easier set of booklets are bigger than for the standard set of booklets.

After scaling the PISA 2009 data for each country separately, the booklet parameters were added to the students' achievement scores for mathematics, reading and science. The mean performance scores could be compared across countries and across booklets. Tables 12.19 to 12.21 present the results of testing the variance in booklet means by country (UH booklet excluded), in each domain. The table rows represent countries and the columns booklets, the cells contain the mean performance by booklet and the square root of the squared difference between the observed and expected mean, divided by the error variance by booklet (a z-score). The expected mean is the average of the booklet means, each weighted by the reciprocal of their error variance. The sum of the squared differences divided by their error variance is chi-square distributed with $13-1=12$ degrees of freedom (where 13 represent the number of booklets). Significant values are in bold.

A z-score is an indication of the magnitude of the difference between the observed booklet mean and the expected booklet mean. Significantly easier or harder than expected booklets are those with z-score >1.96 . Booklets numbers shaded in grey are booklets without items in the domain.

There is no significant booklet effect at the OECD and international level, because the booklet corrections controlled for this effect.

The booklets means for domains that are not included in the booklet (shaded booklets numbers for mathematics and science) do not significantly differ from the expected booklet means for all countries, which is to be expected using the deviation contrast codes for booklets in the conditioning model.

Estimation of the booklet effect for the DRA was not necessary as there were no minor domains included. Table 12.22 presents the results of testing the variance in test form means by country. The TestID 7 column represents imputed scores for the students who did not take the DRA assessment. The chi-square statistics distributed with $7-1=6$ degrees of freedom. There was no significant booklet effect at the international and country level for DRA.

Table 12.22 Variance in DRA booklet means

| | Expected mean | TestID 1 | | TestID 2 | | TestID 3 | | TestID 4 | | TestID 5 | | TestID 6 | | TestID 7 | | Chi-sq (df=6) | |
|-----------------|-----------------|----------|---------|----------|---------|----------|---------|----------|---------|----------|---------|----------|---------|----------|---------|---------------|------|
| | | Mean | Z-Score | Mean | Z-Score | Mean | Z-Score | Mean | Z-Score | Mean | Z-Score | Mean | Z-Score | Mean | Z-Score | | |
| <i>OECD</i> | Australia | 540.2 | 545.0 | 0.9 | 544.4 | 1.0 | 539.3 | 0.2 | 551.2 | 2.1 | 533.9 | 1.1 | 543.0 | 0.6 | 535.0 | 1.9 | 7.5 |
| | Austria | 457.4 | 458.1 | 0.1 | 451.5 | 1.0 | 447.8 | 1.4 | 455.2 | 0.4 | 467.6 | 1.8 | 457.2 | 0.0 | 460.2 | 0.6 | 5.2 |
| | Belgium | 509.9 | 515.9 | 1.4 | 505.9 | 0.9 | 515.1 | 1.1 | 510.7 | 0.2 | 515.1 | 1.3 | 516.1 | 1.6 | 504.3 | 2.4 | 8.9 |
| | Chile | 432.0 | 421.3 | 1.7 | 431.1 | 0.1 | 425.7 | 1.0 | 432.5 | 0.1 | 437.0 | 1.0 | 428.8 | 0.5 | 436.8 | 1.2 | 5.7 |
| | Denmark | 489.6 | 493.0 | 0.6 | 484.5 | 0.8 | 486.5 | 0.4 | 488.0 | 0.2 | 492.6 | 0.4 | 501.7 | 1.8 | 488.2 | 0.6 | 4.8 |
| | France | 478.7 | 482.5 | 0.6 | 473.4 | 0.8 | 476.7 | 0.3 | 488.6 | 1.5 | 482.7 | 0.6 | 483.6 | 0.8 | 472.0 | 1.6 | 6.3 |
| | Hungary | 496.9 | 497.4 | 0.1 | 496.9 | 0.0 | 498.0 | 0.2 | 507.4 | 1.6 | 496.9 | 0.0 | 490.7 | 0.8 | 492.7 | 0.8 | 3.4 |
| | Iceland | 460.8 | 456.6 | 0.6 | 454.5 | 0.9 | 438.8 | 2.4 | 450.1 | 1.3 | 451.3 | 1.4 | 457.6 | 0.5 | 480.5 | 4.3 | 11.4 |
| | Ireland | 508.7 | 508.1 | 0.1 | 501.5 | 1.2 | 505.2 | 0.5 | 510.8 | 0.4 | 513.7 | 0.8 | 509.0 | 0.0 | 509.4 | 0.2 | 3.2 |
| | Japan | 511.7 | 514.9 | 0.4 | 507.5 | 0.6 | 515.0 | 0.4 | 520.4 | 1.0 | 514.6 | 0.4 | 514.3 | 0.4 | 510.9 | 0.5 | 3.7 |
| | Korea | 521.7 | 528.9 | 1.5 | 524.3 | 0.5 | 519.1 | 0.5 | 519.0 | 0.6 | 528.4 | 1.4 | 532.9 | 2.2 | 511.2 | 3.1 | 9.9 |
| | New Zealand | 567.2 | 573.6 | 1.3 | 569.1 | 0.4 | 561.4 | 1.3 | 560.0 | 1.3 | 565.2 | 0.3 | 570.2 | 0.7 | 568.0 | 0.3 | 5.3 |
| | Norway | 500.6 | 504.1 | 0.7 | 490.3 | 2.2 | 513.8 | 2.6 | 511.8 | 2.0 | 500.9 | 0.1 | 495.1 | 1.2 | 498.0 | 0.8 | 9.5 |
| | Poland | 539.8 | 541.2 | 0.2 | 542.6 | 0.5 | 537.6 | 0.4 | 548.1 | 1.2 | 554.5 | 2.9 | 547.0 | 1.3 | 532.5 | 2.8 | 9.2 |
| | Spain | 462.5 | 453.8 | 1.6 | 454.7 | 1.5 | 459.1 | 0.6 | 469.5 | 1.3 | 467.5 | 0.8 | 465.1 | 0.5 | 464.9 | 0.7 | 7.0 |
| Sweden | 513.5 | 507.3 | 1.1 | 519.9 | 1.3 | 512.8 | 0.1 | 519.9 | 1.0 | 516.7 | 0.6 | 519.4 | 1.1 | 506.3 | 2.0 | 7.2 | |
| <i>Partners</i> | Colombia | 369.2 | 372.1 | 0.5 | 359.4 | 1.4 | 362.4 | 0.9 | 373.6 | 0.7 | 377.8 | 1.3 | 370.7 | 0.2 | 368.2 | 0.3 | 5.3 |
| | Hong Kong-China | 514.2 | 515.4 | 0.3 | 517.0 | 0.7 | 507.5 | 1.4 | 505.3 | 1.8 | 514.6 | 0.1 | 520.5 | 1.3 | 515.4 | 0.4 | 5.8 |
| | Macao-China | 492.0 | 488.7 | 0.9 | 488.2 | 1.0 | 483.1 | 2.7 | 489.2 | 0.9 | 494.6 | 0.8 | 488.3 | 1.1 | 494.3 | 2.0 | 9.5 |

Overview of the PISA cognitive reporting scales

PISA 2009 is the fourth PISA assessment and also the fourth occasion on which reading, mathematics and science literacy scores have been reported. A central aim of PISA is to monitor trends over time in indicators based upon reading, mathematics and science literacy. In this section we review the stability of the PISA scales over time, with a view to:

- setting out the range of scales that have been prepared over the past four PISA assessments;
- describing their special features and appropriate use; and
- asking recommendations regarding future design elements of PISA.

Table 12.23 provides a listing of the 19 distinct cognitive scales that have been produced as part of PISA 2000, PISA 2003, PISA 2006 and PISA 2009. For the purpose of this overview, the cognitive scales are classified into three types: PISA literacy scales, PISA literacy subscales and special purpose scales. The PISA literacy scales are the key reporting scales that have been established for each domain, when that domain has been the major domain. The PISA literacy subscales are sub-components of PISA overall literacy scales that were provided when a domain was the major domain. The special purpose scales are additional scales that can be used as interim and trend scales prior to the establishment of the related PISA overall literacy scales.

In the table each scale is named, the database upon which it was established is given, the datasets for which it is provided are indicated (a "P" indicates that the dataset exists); and comments are made about the scale's appropriate use. In the text following, further details are provided on these scales.

Table 12.23 Summary of PISA cognitive reporting scales

| Name | Established | 2000 | 2003 | 2006 | 2009 | Comment |
|---|-------------|------|------|------|------|--|
| PISA literacy scale | | | | | | |
| Print reading | 2000 | P | P | P | P | Trends can be reported between any of the three cycles, by country or by subgroups within countries. |
| Print mathematics | 2003 | | P | P | P | Trends can be reported between 2003, 2006 and 2009 by country or by subgroups within countries. |
| Print science | 2006 | | | P | P | Trends can be reported between 2006 and 2009 by country or by subgroups within countries. |
| PISA literacy subscales | | | | | | |
| Reading subscale: Retrieving Information | 2000 | P | | | P | |
| Reading subscale: Interpreting Texts | 2000 | P | | | P | |
| Reading subscale: Reflection and Evaluation | 2000 | P | | | P | |
| Reading subscale: Continuous Texts | 2009 | | | | P | |
| Reading subscale: Non-Continuous Texts | 2009 | | | | P | |
| Mathematics subscale: Quantity | 2003 | | P | | | |
| Mathematics subscale: Uncertainty | 2003 | | P | | | |
| Mathematics subscale: Space and Shape | 2003 | P | P | | | Established in 2003 and then applied to 2000 with a rescaling (no conditioning). Trends can be reported for countries, but are not optimal for subgroups within countries. |
| Mathematics subscale: Change and Relationships | 2003 | P | P | | | Established in 2003 and then applied to 2000 with a rescaling (no conditioning). Trends can be reported for countries, but are not optimal for subgroups within countries. |
| Science subscale: Explaining Phenomena Scientifically | 2006 | | | P | | |
| Science subscale: Identifying Scientific Issues | 2006 | | | P | | |
| Science subscale: Using Scientific Evidence | 2006 | | | P | | |
| Science subscale: Physical Systems | 2006 | | | P | | Limited conditioning implemented permitting unbiased estimation by country and by gender. Results for other subgroups are not optimal. |
| Science subscale: Earth and Space Systems | 2006 | | | P | | Limited conditioning implemented permitting unbiased estimation by country and by gender. Results for other subgroups are not optimal. |
| Science subscale: Living Systems | 2006 | | | P | | Limited conditioning implemented permitting unbiased estimation by country and by gender. Results for other subgroups are not optimal. |
| Special purpose scales | | | | | | |
| Interim mathematics | 2000 | P | | | | |
| Interim science | 2000 | P | P | | | |
| Science trend 2003-2006 | 2006 | | P | P | | Uses items that were common to PISA 2003 and 2006. |
| Electronic reading | 2009 | | | | P | |

PISA literacy scales

The primary PISA reporting scales are reading, mathematics and science. These scales were established in the year in which the respective domain was the major domain, since in that year the framework for the domain was fully developed and the domain was comprehensively assessed. When the overall literacy scale is established the mean of the scale is set at 500 and the standard deviation is set at 100 (for the pooled, equally weighted OECD countries) – for example, 500 on the PISA mathematics scale is the mean achievement of assessed students in OECD countries in 2003.

The intention is that these overall literacy scales will stay in place until the specification of the domain is changed or updated.

PISA literacy subscales

Across the four PISA assessments a total of 19 subscales have been prepared and reported. In PISA 2000, three reading aspect-based scales were prepared; in PISA 2003, four mathematics content-based scales were prepared, in 2006 a total of six science scales were prepared; and in PISA 2009 two text format scales were prepared.

The subscales are typically prepared only in the year in which a domain is a major domain, since when a domain is a major domain there are sufficient items in each sub-area to support the reporting of the scales. The one exception to this general practice is mathematics, for which the *space and shape* and *change and relationships* scales were reported for the PISA 2000 data as well as the PISA 2003 data. These scales, which were established in 2003 when mathematics was the major domain, could be applied to the 2000 data because only these two areas of mathematics had been assessed in PISA 2000 and sufficient common items were available to support the scaling.



For the 2000 data the mathematics scales were prepared using a methodology that permits trend analysis at the national level (or at the level of adjudicated regions), but the scales are not optimal for analysis at the level of student sub-groups.³

For science in PISA 2006, two alternative sets of scales were prepared. The first was a set of three process-based scales and the second was a set of three content-based scales. It is important to note that these are alternative scalings that each rely on the same test items. As such, it is inappropriate to jointly analyse scales that are selected from the alternative scalings. For example, it would not be meaningful or defensible to correlate or otherwise compare performance on the “Physical systems” scale, with performance on the *using scientific evidence* scale. Furthermore the content-based scales can be analysed at the national level (or at the level of adjudicated regions), and can be analysed by gender, but they are not optimal for use at the level of any other student sub-groups, whereas the process-based scales are suitable in addition for sub-group analyses.⁴

The metric of all of the PISA subscales is set so that scales within a domain can be compared to each other and with the matching overall PISA reporting scale.⁵

Special purpose scales

There are three special purpose scales.

An interim mathematics scale was established and reported in PISA 2000. This scale was prepared to provide an overall mathematics score, and it used all of the mathematics items that were included in the PISA 2000 assessment. This scale was discontinued in 2003 when mathematics was the major domain and the alternative and more comprehensive PISA overall mathematics literacy scale was established.

An interim science scale was established and reported in PISA 2000. This scale was prepared to provide an overall science score, and it used all of the science items that were included in the PISA 2000 assessment. The PISA 2003 science data were linked to this scale so that the PISA 2003 science results were also reported on this interim science scale. For PISA 2006 this scale was not provided since science was the major domain and the alternative and more comprehensive overall PISA science scale was established.

To allow comparisons between science outcomes in 2003 and 2006 a science trend 2003-2006 scale was prepared. This scale is based upon the science items that are common to PISA 2003 and 2006 and can be used to examine trends (on those common items) between 2003 and 2006. The PISA 2003 abilities that are based on the common items can be analysed at the national level (or at the level of adjudicated regions), and can be analysed by gender, but they are not optimal for use at the level of any other student sub-groups. The PISA 2006 abilities, associated with the fully developed overall PISA science scale, can be analysed by national subgroups as well.

OBSERVATIONS CONCERNING THE CONSTRUCTION OF THE PISA OVERALL LITERACY SCALES

A number of the PISA scales have been established to permit trend analyses. A review of the various links available and necessary to establish these scales is given below. Table 12.24 illustrates the nine linkages of the PISA domains that are examined and discussed below. Links (1), (2) and (3) are for reading 2000 to 2003, 2003 to 2006 and 2006 to 2009 respectively, links (4), (5) and (6) are for mathematics 2000 to 2003, 2003 to 2006 and 2006 to 2009 respectively, links (7), (8) and (9) are for science 2000 to 2003, 2003 to 2006 and 2006 to 2009 respectively.

Table 12.24 also indicates in which data collections the domain was a major domain and on which occasions it was a minor domain. As a consequence one can note that on three occasions the links are major to minor (links (1), (5) and (9)), on three occasions they are minor to minor (links (2), (6) and (7)), and on three occasions they are minor to major (links (3), (4) and (8)).

Table 12.24 Linkage types among PISA domains 2000 - 2009

| | 2000 | | 2003 | | 2006 | | 2009 |
|-------------|-------|----------|-------|----------|-------|----------|-------|
| | Major | (1) → | Minor | (2) → | Minor | (3) → | Major |
| Reading | Major | (1) → | Minor | (2) → | Minor | (3) → | Major |
| Mathematics | Minor | (4) → | Major | (5) → | Minor | (6) → | Minor |
| Science | Minor | (7) → | Minor | (8) → | Major | (9) → | Minor |

When a proficiency area is assessed as a major domain there are two key characteristics that distinguish it from a minor domain. First the framework for the area is fully developed and elaborated. Second the framework is comprehensively assessed since more assessment time is allocated to the major domain than is allocated to each of the minor domains.

Framework development

For PISA 2000 a full and comprehensive framework was developed for reading to guide the assessment of reading as a major domain. Less fully articulated frameworks were developed to support the assessment of mathematics and science as minor domains.⁶

For PISA 2003, the mathematics framework was updated and fully developed to support a comprehensive assessment of mathematics. The science frameworks were retained largely as they had been for PISA 2000.⁷

The key changes to the mathematics framework between 2000 and 2003 were:

- addition of a theoretical underpinning of the mathematics assessment, expanding the rationale for the PISA emphasis on using mathematical knowledge and skills to solve problems encountered in life;
- restructuring and expansion of domain content: expansion from two broad content areas (overarching ideas) to four; removal of all reference to mathematics curricular strands as a separate content categorisation (instead, definitions of the overarching ideas were expanded to include the kinds of school mathematics topics associated with each);
- a more elaborated rationale for the existing balance between realistic mathematics and more traditional context-free items, in line with the literacy for life notion underlying OECD/PISA assessments;
- a redeveloped discussion of the relevant mathematical processes: a clearer and much enhanced link between the process referred to as mathematisation, the underlying mathematical competencies, and the competency clusters; and a better operationalisation of the competency classes through a more detailed description of the underlying proficiency demands they place on students; and
- considerable elaboration through addition of examples, including items from previous test administrations.

Clearly, the framework change involving an effective doubling of the mathematical content base of the study was of such significance that trend measures would be very seriously affected. Hence, only scale links to 2000 were possible, and the new framework provided the first comprehensive basis for the calculation of future trend estimates.

For PISA 2006, science was the major domain so the science framework was updated and fully developed to support a comprehensive assessment of science. The reading framework was retained largely as it had been for PISA 2000, and the mathematics framework as it had been for PISA 2003.⁸ The key changes to the science framework between 2003 and 2006 as they relate to comparison in the science scales over time were:

- A clearer separation of knowledge about science as a form of human enquiry from knowledge of science, meaning knowledge of the natural world as articulated in the different scientific disciplines. In particular, PISA 2006 gives greater emphasis to knowledge about science as an aspect of science performance, through the addition of elements that underscore students' knowledge about the characteristic features of science and scientific endeavour.
- The addition of new components on the relationship between science and technology.

Both of these changes carry the potential to disrupt links with the previous special purpose science scales: the interim science and trend science scales.

With regard to reading, much of the substance of the PISA 2000 framework was retained in the PISA 2009 framework, respecting one of the central purposes of the PISA project: to collect and report trend information about performance in reading, mathematics and science. However, the PISA domain frameworks are also aimed to be evolving documents that will adapt to and integrate new developments in theory and practice over time. There was therefore some evolution, reflecting both an expansion in our understanding of the nature of reading and changes in the world. At the same time there was no need to develop a new scale for reading, so that performance from 2009 could be compared to 2000.

There were two major modifications in the reading framework:

- incorporating the reading of electronic texts; and
- elaborating the constructs of reading engagement and metacognition.



Testing time and item characteristics

In each of PISA 2000, PISA 2003 and PISA 2006 a total of 390 minutes of testing material was used.⁹ In this case there were thirteen 30 minutes clusters of items (390 minutes all together). These 13 clusters were included in 13 two-hour booklets (4 clusters in each booklet). In PISA 2009, due to the addition of the easy booklets, a total of 450 minutes of testing material was used.¹⁰

The distribution of the testing minutes is given in Table 12.25. When a domain is assessed as a major domain then more minutes are devoted to it than for minor domains. For example 270 minutes were assigned to reading material in PISA 2000 and PISA 2009 to allow full coverage of the framework. Similarly, PISA 2003 included 210 minutes of mathematics material and PISA 2006 included 210 minutes of science material. When a domain is assessed as a minor domain the assessment is less comprehensive and does not provide an in-depth assessment of the full framework that is developed when a domain is a major domain.

It is also important to recognise that given the PISA test design (see Chapter 2) the change of major domains over time means that the testing experience for the majority of students will be different in each cycle because it becomes dominated by the new major domain. For example, the design for PISA 2009 used 13 booklets per country. Ten of them comprised at least 50% of reading material. For three of these the other 50% comprised only mathematics material, three were completed with a mixture of science and mathematics material, other three were completed with the mixture of reading and science. One booklet contained only reading material. Remaining three booklets contained one reading, one mathematics and two science clusters.

This could be compared to the design for PISA 2006 that also used 13 booklets. Eleven of them comprised at least 50% of science material. For four of these the other 50% comprised only mathematics material, four were completed with a mixture of reading and mathematics material, and for one booklet the other 50% comprised only reading material. Two booklets contained only science material.

The links in terms of numbers of items in common for successive pairs of assessments are shown in Table 12.26.

Table 12.25 Number of unique item minutes for each domain for each PISA assessments

| | Reading | Mathematics | Science | Total |
|------|---------|-------------|---------|------------------|
| 2000 | 270 | 60 | 60 | 390 |
| 2003 | 60 | 210 | 60 | 330 ¹ |
| 2006 | 60 | 120 | 210 | 390 |
| 2009 | 270 | 90 | 90 | 450 ² |

1. 60 minutes were devoted to problem solving.

2. 390 minutes unique item minutes per country.

Characteristics of each of the links

To allow a comparison between PISA cycles a set of the same items (link items) included for each domain in each PISA assessment. The number of link items in each domains included in Table 12.26.

Table 12.26 Numbers of link items between successive PISA assessments*

| | Reading | Mathematics | Science |
|----------------|---------|-------------|---------|
| Link 2000-2003 | 28 | 20 | 25 |
| Link 2003-2006 | 28 | 48 | 22 |
| Link 2000-2009 | 26 | 8 | 5 |
| Link 2003-2009 | 26 | 35 | 9 |
| Link 2006-2009 | 26 | 35 | 53 |

* Total number of items included in major domains Reading 2000, Mathematics 2003, Science 2006 and Reading 2009 are 129, 84, 108 and 131 respectively.

Reading 2000 to 2003

The PISA reading scale was established in 2000 on the basis of a fully developed and articulated framework and a comprehensive assessment of that framework. The PISA 2000 included 129 reading items. In PISA 2003 a subset of 28 of the 2000 reading items was selected and used. Equating procedures reported in OECD (2005) were then used to report the PISA 2003 data on the established PISA reading scale.

The trend results for the OECD countries that participated in both PISA 2000 and PISA 2003 showed that of 32 countries, 10 had a significant decline in mean score and 5 had a significant rise in mean score (OECD, 2004).

When reviewing the potential causes for this possible instability a number of relevant issues were observed. First, there was a substantial test design change between PISA 2000 and PISA 2003. The PISA 2003 design was fully balanced whereas the PISA 2000 design systematically placed minor domain items and some reading items at the end of the student booklets (see Adams and Wu, 2002). The complexity of the PISA 2000 design is such that the impact of this on the item parameter estimation and hence the equating is unclear. Second, the units that were selected from PISA 2000 for use in PISA 2003 were edited in minor ways. While none of the individual link items was edited, some items in the units were removed. As with the test design change, the impact of this change on the item parameter estimation and hence the equating is unclear. Third, the clusters of items that were used were not pre-existing clusters. In particular, units from PISA 2000 clusters one to seven were selected and reconstituted as two new clusters. Intact clusters of items could not be used from PISA 2000 since none of the individual pre-existing clusters provided an adequate coverage of the framework.

The percentage correct on reading items that link PISA 2000 and PISA 2003 are given in Table 12.27, with the corresponding scatter plot in Figure 12.5. To compute the percentage correct, all students were included from countries that were included in trend analysis between PISA 2000 and PISA 2003. For this analysis 25 OECD countries were included. Those excluded were the United Kingdom (who did not meet school response rate in 2003), the Netherlands (who did not meet school response rate in 2000), Luxembourg (who used multilingual booklets in 2000), and the Slovak Republic and Turkey (who did not participate in the PISA 2000 study). In addition, recent OECD members such as Chile, Estonia, Israel and Slovenia were not included.

The mean of the differences between PISA 2003 and PISA 2000 is -1.11, and the standard deviation of the differences is 2.82.

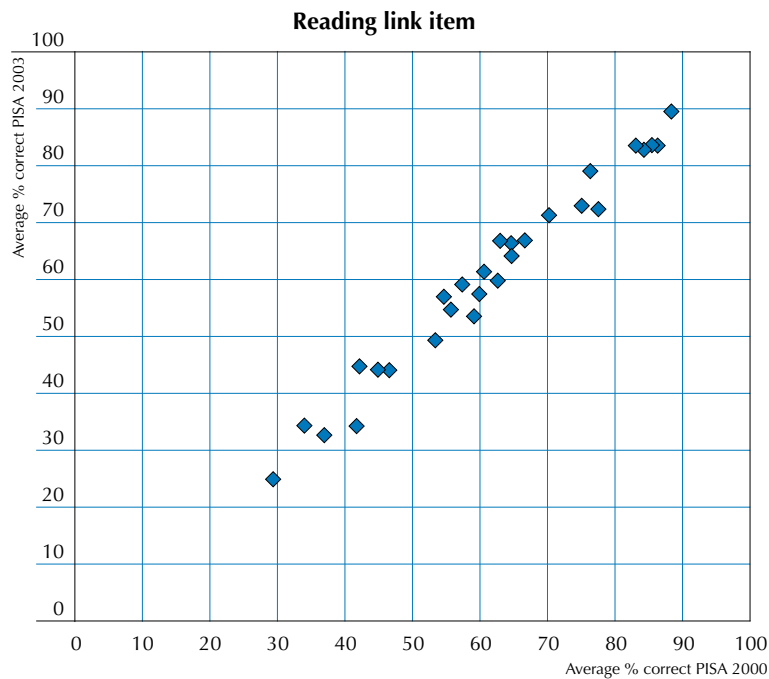
Table 12.27 International percent correct for reading link items in PISA 2000 and PISA 2003

| Item | % correct | |
|----------|-----------|------|
| | 2000 | 2003 |
| R055Q01 | 84.4 | 82.9 |
| R055Q02 | 53.4 | 49.1 |
| R055Q03 | 62.7 | 59.8 |
| R055Q05 | 77.7 | 72.5 |
| R067Q01 | 88.5 | 89.7 |
| R067Q04 | 54.7 | 57 |
| R067Q05 | 62.9 | 67.1 |
| R102Q04A | 37.1 | 32.4 |
| R102Q05 | 42.2 | 44.9 |
| R102Q07 | 86.2 | 83.5 |
| R104Q01 | 83 | 83.2 |
| R104Q02 | 41.6 | 34.5 |
| R104Q05 | 29.2 | 24.9 |
| R111Q01 | 64.8 | 66.3 |
| R111Q02B | 34.2 | 34 |
| R111Q06B | 44.8 | 44.5 |
| R219Q01 | 70.2 | 71.2 |
| R219Q01E | 57.4 | 59.3 |
| R219Q02 | 76.5 | 78.8 |
| R220Q01 | 46.8 | 44.4 |
| R220Q02B | 64.8 | 64 |
| R220Q04 | 60.8 | 61.3 |
| R220Q05 | 85.5 | 83.2 |
| R220Q06 | 66.6 | 67.1 |
| R227Q01 | 59 | 53.8 |
| R227Q02 | 59.8 | 57.7 |
| R227Q03 | 56 | 54.9 |
| R227Q06 | 75.2 | 72.9 |



■ Figure 12.5 ■

Scatter plot of percentage correct for reading link items in PISA 2000 and PISA 2003



Reading 2003 to 2006

To link the PISA 2006 data to the PISA reading scale the same 28 items (units and clusters) as were used in PISA 2003 were again used. The trend results for the OECD countries that participated in both PISA 2003 and PISA 2006 showed that of the 38 countries which could be compared, five had a significant decline in mean score and two had a significant rise in mean score (OECD, 2007). The number of significant changes was less than reported for the 2000-2003 link.

A number of reasons might be conjectured as possible explanations of this lack of consistency. First, presenting a large number of reading items with a small number of mathematics and science items interspersed, provides for a very different test-taking experience for students compared to a test with a majority of mathematics items, and a few reading, general problem solving and science items interspersed. This may have impacted on the trend estimates. Second, the mix of reading items by aspect type was somewhat different between the two test administrations. In 2003 there was a larger proportion of score points in the reflection and evaluation aspect than had been the case for 2006.

The percentage correct on reading items that link PISA 2003 and PISA 2006 are given in Table 12.28, with the corresponding scatter plot in Figure 12.6. To compute the percentage correct, all students were included from countries that were included in these trend analyses. For percentage correct, 28 OECD countries were included. Excluded were the United Kingdom (because of low response rate) and the United States (reading scores are not available for PISA 2006 because of a printing error).

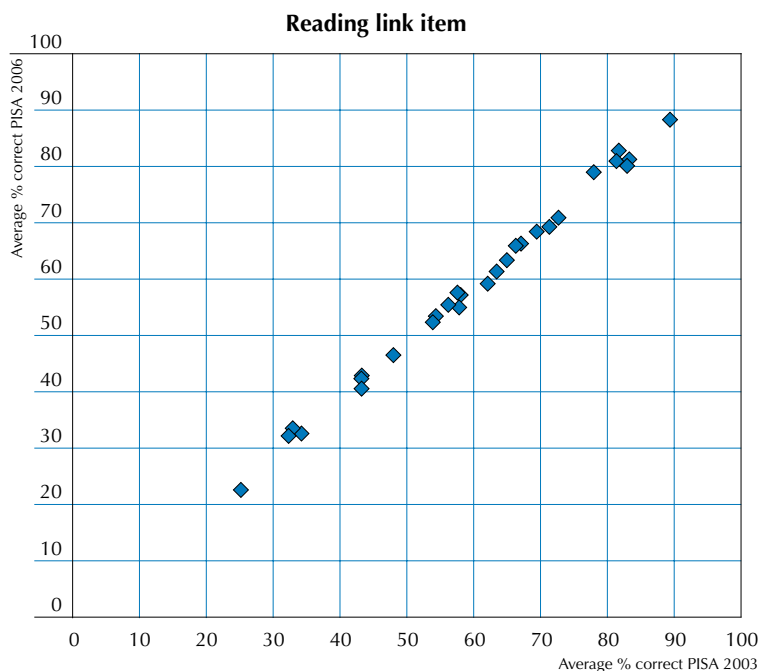
The mean of the differences between PISA 2003 and PISA 2006 is -1.17 (PISA 2006 minus PISA 2003), and the standard deviation of the differences is 1.07. The standard deviation of this difference is much less than that for 2000 to 2003 and most likely due to the use of identical items in identical clusters for the two assessments.

Table 12.28 International percent correct for reading link items in PISA 2003 and PISA 2006

| Item | % correct | |
|----------|-----------|------|
| | 2003 | 2006 |
| R055Q01 | 81.4 | 80.9 |
| R055Q02 | 47.9 | 46.8 |
| R055Q03 | 58.2 | 57.2 |
| R055Q05 | 72.6 | 71 |
| R067Q01 | 89.5 | 88.2 |
| R067Q04 | 56.1 | 55.6 |
| R067Q05 | 66.4 | 65.9 |
| R102Q04A | 32.4 | 32.2 |
| R102Q05 | 43.1 | 42.8 |
| R102Q07 | 81.8 | 82.9 |
| R104Q01 | 83 | 80.3 |
| R104Q02 | 34.3 | 32.9 |
| R104Q05 | 25.3 | 22.8 |
| R111Q01 | 64.9 | 63.4 |
| R111Q02B | 32.9 | 33.4 |
| R111Q06B | 43.3 | 40.9 |
| R219Q01 | 69.6 | 68.4 |
| R219Q01E | 57.5 | 57.4 |
| R219Q02 | 78.1 | 78.8 |
| R220Q01 | 43.2 | 42.5 |
| R220Q02B | 63.5 | 61.2 |
| R220Q04 | 62.1 | 59.2 |
| R220Q05 | 83.2 | 81 |
| R220Q06 | 67.1 | 66.4 |
| R227Q01 | 53.7 | 52.3 |
| R227Q02 | 57.9 | 55 |
| R227Q03 | 54.4 | 53.3 |
| R227Q06 | 71.3 | 69.3 |

■ Figure 12.6 ■

Scatter plot of percentage correct for reading link items in PISA 2003 and PISA 2006

**Reading 2000 to 2009**

To link the PISA 2009 data to the PISA reading scale the same 28 items (units and clusters) used in both PISA 2003 and PISA 2006 were again used. Two link items were deleted from the link item set because of data entry errors. The trend results for the OECD countries that participated in both PISA 2000 and PISA 2009 showed that of the 38 countries which could be compared, 5 had a significant decline in mean score and 13 had a significant rise in mean score (OECD, 2010b).



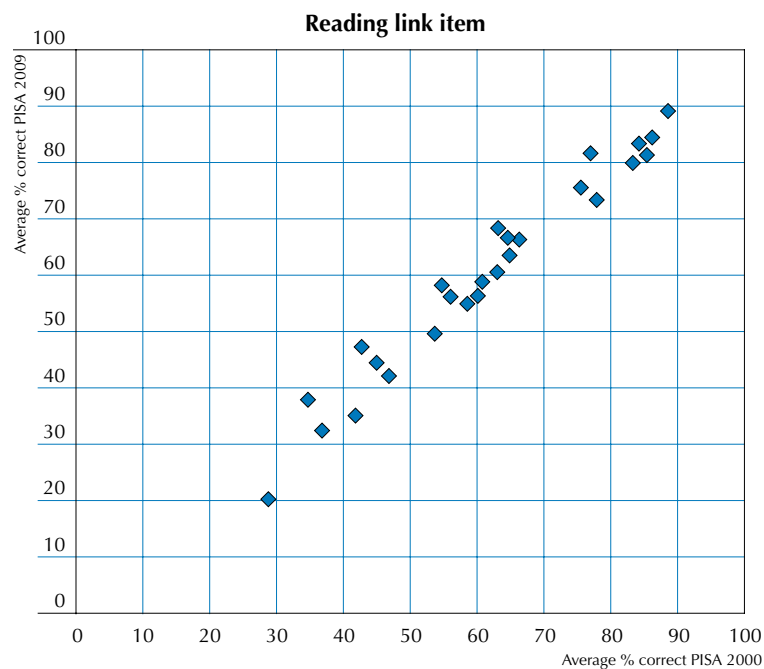
The percentage correct on reading items that link PISA 2000 and PISA 2009 are given in Table 12.29, with the corresponding scatter plot in Figure 12.7. To compute the percentage correct, all students were included from countries that were included in these trend analyses. For percentage correct, 26 OECD countries were included. Excluded were the Netherlands, Luxembourg, the Slovak Republic and Turkey. The mean of the differences (PISA 2000 minus PISA 2009) is 1.30, and the standard deviation of the differences is 3.53. The standard deviation of this difference is greater than that for 2003 to 2006 but comparable to the 2000 to 2003 difference and most likely due to the inclusion of the new item clusters.

Table 12.29 International percent correct for reading link items in PISA 2000 and PISA 2009

| Item | % correct | |
|----------|-----------|------|
| | 2000 | 2009 |
| R055Q01 | 84.4 | 82.9 |
| R055Q02 | 53.5 | 49.7 |
| R055Q03 | 63.0 | 60.7 |
| R055Q05 | 77.8 | 73.5 |
| R067Q01 | 88.5 | 89.2 |
| R067Q04 | 55.0 | 57.9 |
| R067Q05 | 63.2 | 68.3 |
| R102Q04A | 36.9 | 32.5 |
| R102Q05 | 42.9 | 47.2 |
| R102Q07 | 86.3 | 84.5 |
| R104Q01 | 83.2 | 80.2 |
| R104Q02 | 41.8 | 34.9 |
| R104Q05 | 29.1 | 20.3 |
| R111Q01 | 64.7 | 66.7 |
| R111Q02B | 34.8 | 37.9 |
| R111Q06B | 45.1 | 44.4 |
| R219Q02 | 77.1 | 81.6 |
| R220Q01 | 46.7 | 41.9 |
| R220Q02B | 64.9 | 63.6 |
| R220Q04 | 60.6 | 58.6 |
| R220Q05 | 85.4 | 81.4 |
| R220Q06 | 66.2 | 66.6 |
| R227Q01 | 58.6 | 55.1 |
| R227Q02 | 59.9 | 56.3 |
| R227Q03 | 56.2 | 56.1 |
| R227Q06 | 75.4 | 75.5 |

■ Figure 12.7 ■

Scatter plot of percentage correct for reading link items in PISA 2000 and PISA 2009



Mathematics 2000 to 2003

The mathematics framework that was prepared for PISA 2000 was preliminary and the assessment was restricted to two of the so-called big ideas – *space and shape*, and *change and relationships*. For the PISA 2003 assessment, when mathematics was a major domain, the framework was fully developed and the assessment was broadened to cover the four overarching ideas – *quantity*, *uncertainty*, *space and shape*, and *change and relationships*.

Given that the mathematics framework was fully developed for PISA 2003, the PISA mathematics scale was developed at that point. As PISA 2000 had covered two of the four 2003 mathematics scales, only two trend scales could be developed. These were for comparison of performance between 2000 and 2003 for *space and shape*, and *change and relationships*.

PISA 2000 and PISA 2003 percentages correct for mathematics *space and shape* and *change and relationships* link items are given in Table 12.28, with the corresponding scatter plot in Figure 12.8. Similar to the reading 2000 to 2003 item analysis student responses from only 25 OECD countries were included in computation of the percentage correct.

The mean of the differences between PISA 2003 and PISA 2000 is 2.39, and the standard deviation of the differences is 2.79.

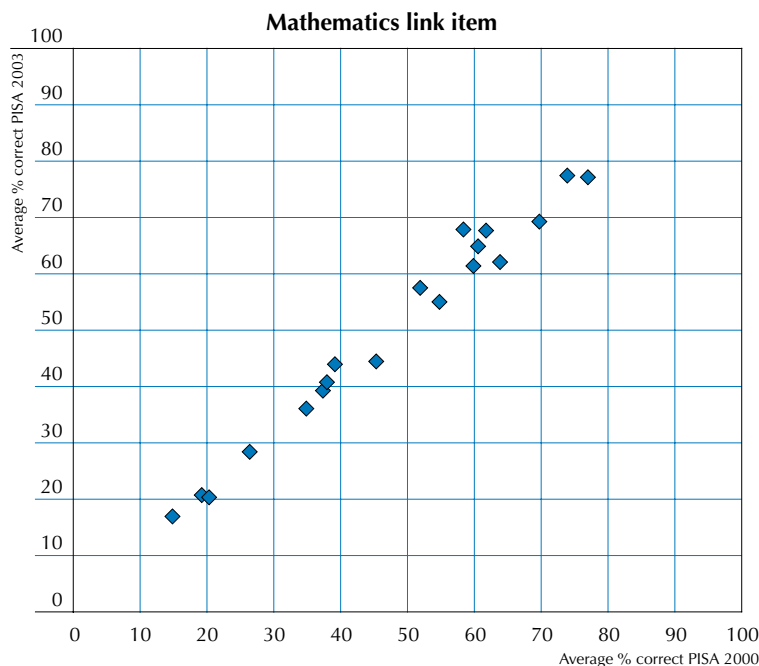
Table 12.30 International percent correct for mathematics link items in PISA 2000 and PISA 2003

| Item | % correct | |
|---------|-----------|------|
| | 2000 | 2003 |
| M033Q01 | 74.2 | 77.8 |
| M034Q01 | 39.3 | 44.2 |
| M124Q01 | 35.0 | 36.3 |
| M124Q03 | 19.3 | 20.9 |
| M144Q01 | 64.1 | 62.4 |
| M144Q02 | 26.5 | 28.6 |
| M144Q03 | 77.3 | 77.5 |
| M144Q04 | 37.5 | 39.5 |
| M145Q01 | 58.6 | 68.2 |
| M150Q01 | 62.0 | 68.0 |
| M150Q02 | 70.0 | 69.6 |
| M150Q03 | 45.5 | 44.7 |
| M155Q01 | 60.8 | 65.2 |
| M155Q02 | 60.1 | 61.7 |
| M155Q03 | 14.9 | 17.1 |
| M155Q04 | 52.1 | 57.8 |
| M192Q01 | 38.1 | 41.0 |
| M266Q01 | 20.4 | 20.5 |
| M273Q01 | 55.0 | 55.3 |



■ Figure 12.8 ■

Scatter plot of percentage correct for mathematics space and shape and change and relationships link items in PISA 2000 and PISA 2003



Mathematics 2003 to 2006

A set of 48 mathematics items was selected from PISA 2003 and used again in PISA 2006.¹¹ Hence the change from 2003 to 2006 involved reducing the number of items by almost half, and as was the case when reading changed from a major to a minor domain, it was not possible to make such a reduction whilst retaining intact clusters. Four new clusters were formed for PISA 2006 from the units retained from PISA 2003. The trend results for the OECD countries that participated in both PISA 2003 and 2006 showed that of the 39 countries which could be compared 4 had a significant decline in mean score and 4 had a significant rise in mean score (OECD, 2007). The magnitude and number of these changes is consistent with the figures for reading from 2003 to 2006 and with figures observed in other international studies such as TIMSS (Mullis, Martin, and Foy [with Olson, Preuschoff, Erberber, Arora, and Galia], 2008).

The percentage correct on mathematics items that link PISA 2003 and PISA 2006 are given in Table 12.31, with the corresponding scatter plot in Figure 12.9. To compute the percentage correct, all students were included from countries that were included in these trend analyses. For percentage correct, 29 OECD countries were included. The United Kingdom was excluded because it was excluded from PISA 2003.

It is interesting to contrast these results with those observed for reading. At the item level the consistency seems somewhat less for mathematics than for reading, whereas at the scale level the consistency is comparable. It is possible that the item-level inconsistency is caused by the change from mathematics as a major domain to mathematics as a minor domain. Two specific aspects of the change are likely to have contributed to this inconsistency. One is the fact that it was necessary to select a subset of items and form new trend clusters. The rearrangement of items into new clusters appears to have a small impact on relative item difficulty. The second is the fact that the items were presented to students in a different context from before; specifically that the items were no longer from the dominant domain, rather they represented a smaller set of items presented amongst a much larger number of science items.

The mean of the differences (PISA 2003 minus PISA 2006) is 1.40, and the standard deviation of the differences is 1.77. This standard deviation is less than that for reading between 2000 and 2003 but greater than that for reading between 2003 and 2006. This is consistent with the fact that 2003 and 2006 designs were both balanced but, unlike the reading items, the mathematics link items between 2003 and 2006 were not presented in the same clusters.

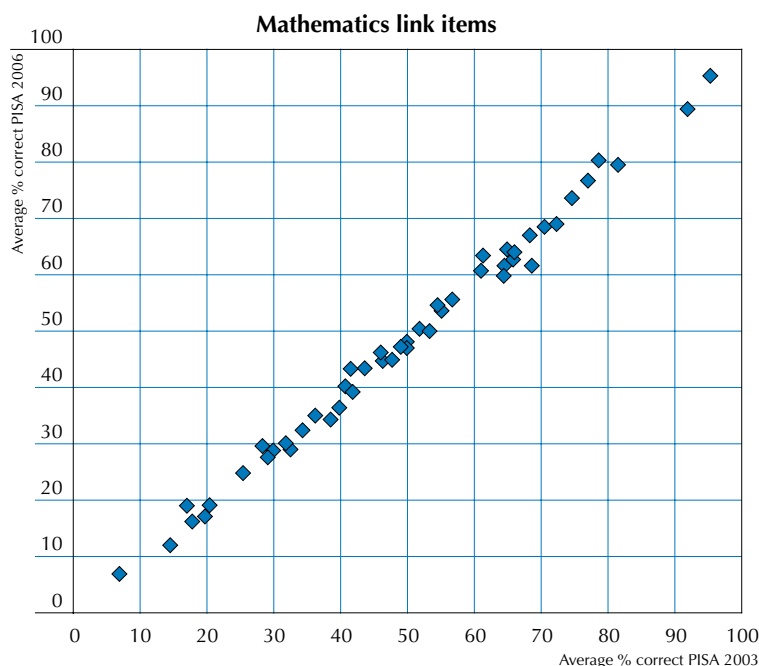
Table 12.31 International percent correct for mathematics link items in PISA 2003 and PISA 2006

| Item | % correct | |
|---------|-----------|------|
| | 2003 | 2006 |
| M033Q01 | 77.0 | 76.8 |
| M034Q01 | 43.6 | 43.5 |
| M155Q01 | 64.9 | 64.6 |
| M155Q02 | 61.0 | 60.8 |
| M155Q03 | 17.0 | 19.1 |
| M155Q04 | 56.7 | 55.7 |
| M192Q01 | 40.7 | 40.3 |
| M273Q01 | 55.1 | 53.7 |
| M302Q01 | 95.3 | 95.4 |
| M302Q02 | 78.6 | 80.4 |
| M302Q03 | 29.9 | 28.9 |
| M305Q01 | 64.5 | 61.7 |
| M406Q01 | 29.1 | 27.7 |
| M406Q02 | 19.7 | 17.2 |
| M408Q01 | 41.5 | 43.4 |
| M411Q01 | 51.8 | 50.5 |
| M411Q02 | 46.3 | 44.8 |
| M420Q01 | 49.9 | 48.2 |
| M421Q01 | 65.8 | 62.8 |
| M421Q02 | 17.8 | 16.3 |
| M421Q03 | 38.5 | 34.4 |
| M423Q01 | 81.5 | 79.6 |
| M442Q02 | 41.8 | 39.3 |
| M446Q01 | 68.3 | 67.1 |
| M446Q02 | 6.9 | 7.0 |
| M447Q01 | 70.5 | 68.6 |
| M462Q01 | 14.5 | 12.1 |
| M464Q01 | 25.4 | 24.9 |
| M474Q01 | 74.6 | 73.7 |
| M496Q01 | 53.3 | 50.1 |
| M496Q02 | 66.0 | 64.1 |
| M559Q01 | 61.3 | 63.5 |
| M564Q01 | 49.9 | 47.1 |
| M564Q02 | 46.0 | 46.3 |
| M571Q01 | 49.0 | 47.3 |
| M598Q01 | 64.4 | 59.9 |
| M603Q01 | 47.7 | 45.0 |
| M603Q02 | 36.2 | 35.1 |
| M710Q01 | 34.3 | 32.5 |
| M800Q01 | 91.9 | 89.5 |
| M803Q01 | 28.3 | 29.7 |
| M810Q01 | 68.6 | 61.7 |
| M810Q02 | 72.3 | 69.1 |
| M810Q03 | 20.4 | 19.2 |
| M828Q01 | 39.8 | 36.5 |
| M828Q02 | 54.5 | 54.7 |
| M828Q03 | 32.5 | 29.1 |
| M833Q01 | 31.8 | 30.2 |



■ Figure 12.9 ■

Scatter plot of percentage correct for mathematics link items in PISA 2003 and PISA 2006



Mathematics 2006 to 2009

A set of 35 mathematics items (three out of four PISA 2006 mathematics clusters) was selected from PISA 2006 and used again in PISA 2009.¹² The trend results for the OECD countries that participated in both PISA 2006 and PISA 2009 showed that of the 55 countries which could be compared 9 had a significant decline in mean score and 11 had a significant rise in mean score (OECD, 2010b).

The percentage correct on mathematics items that link PISA 2006 and PISA 2009 are given in Table 12.32, with the corresponding scatter plot in Figure 12.10. To compute the percentage correct, all students were included from countries that were included in these trend analyses. For percentage correct, 34 OECD countries were included.

The mean of the differences (PISA 2009 minus PISA 2006) is 0.22, and the standard deviation of the differences is 1.36. The standard deviation of this difference is less than that for 2003 to 2006 and most likely due to the use of identical clusters for the two assessments as it was a case for the reading for 2003 to 2006.

[Part 1/2]

Table 12.32 International percent correct for mathematics link items in PISA 2006 and PISA 2009

| Item | % correct | |
|----------|-----------|------|
| | 2006 | 2009 |
| M033Q01 | 76.1 | 75.3 |
| M034Q01T | 42.6 | 42.4 |
| M155Q01 | 64.3 | 66.3 |
| M155Q02D | 59.8 | 61.5 |
| M155Q03D | 18.3 | 18.5 |
| M155Q04T | 55.0 | 54.9 |
| M192Q01T | 39.4 | 41.1 |
| M273Q01T | 52.9 | 52.7 |
| M406Q01 | 26.6 | 26.7 |
| M406Q02 | 16.3 | 16.7 |
| M408Q01T | 42.2 | 40.2 |
| M411Q01 | 49.0 | 47.9 |
| M411Q02 | 44.3 | 44.8 |
| M420Q01T | 47.4 | 50.6 |
| M423Q01 | 79.3 | 79.1 |
| M442Q02 | 38.4 | 38.4 |
| M446Q01 | 66.8 | 69.0 |
| M446Q02 | 6.7 | 7.1 |

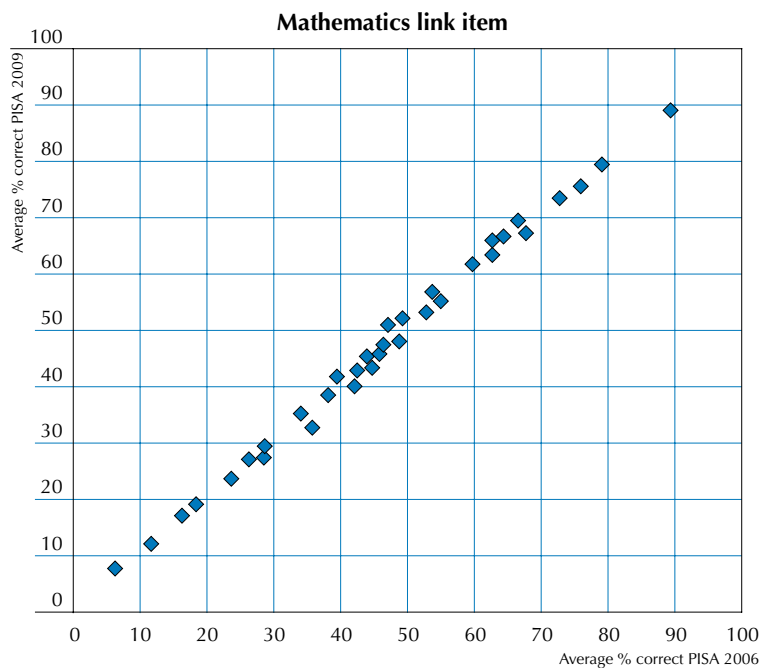
[Part 2/2]

Table 12.32 International percent correct for mathematics link items in PISA 2006 and PISA 2009

| Item | % correct | |
|----------|-----------|------|
| | 2006 | 2009 |
| M447Q01 | 67.6 | 67.4 |
| M462Q01D | 12.0 | 11.4 |
| M464Q01T | 23.8 | 23.2 |
| M474Q01 | 72.9 | 73.1 |
| M496Q01T | 49.4 | 51.5 |
| M496Q02 | 63.1 | 65.7 |
| M559Q01 | 62.8 | 63.3 |
| M564Q01 | 46.5 | 46.4 |
| M564Q02 | 45.9 | 45.8 |
| M571Q01 | 46.5 | 46.6 |
| M603Q01T | 44.5 | 43.5 |
| M603Q02T | 34.3 | 34.8 |
| M800Q01 | 89.3 | 89.0 |
| M803Q01T | 28.7 | 27.3 |
| M828Q01 | 35.9 | 32.3 |
| M828Q02 | 53.9 | 56.0 |
| M828Q03 | 28.8 | 28.5 |

■ Figure 12.10 ■

Scatter plot of percentage correct for mathematics link items in PISA 2006 and PISA 2009

**Science 2000 to 2003**

Science was a minor domain in both PISA 2000 and 2003. As such the assessment on both of these occasions was less comprehensive than it was in 2006, when a more fully articulated framework and more testing time was available. There were 25 items that were common to both PISA 2000 and PISA 2003. The trend results for the OECD countries that had participated in both PISA 2000 and PISA 2003 showed that of 32 countries, 5 had a significant decline in mean score and 13 a significant rise in mean score (OECD, 2004).



The number of inconsistencies between 2000 and 2003 was greater than expected at both the item-level and at the scale level. When reviewing the potential causes for this possible instability a number of relevant issues were observed. First, as mentioned above for reading, there was a substantial test design change between PISA 2000 and PISA 2003. The complexity of the PISA 2000 design is such that impact of this on the item parameter estimation and hence the equating is unclear. Second, the units that were selected from PISA 2000 for use in PISA 2003 were edited in minor ways. As with reading, while none of the link items were edited, some items in the units were removed. And as with the test design change, the impact of this on the item parameter estimation and hence the equating is unclear. Third, the clusters of items that were used were not pre-existing clusters. The material retained from the two PISA 2000 clusters was supplemented with a small number of new units, and reconstituted as two new clusters. Fourth, there were just 25 link items between these two assessments, and unlike mathematics these items were spread across all aspects of the framework. This number was less than desirable and was a result of choices made concerning the release of items following the 2000 assessment to illustrate the nature of the PISA assessment to the public.

The percentage correct on science items that link PISA 2000 and PISA 2003 are given in Table 12.33, with the corresponding scatter plot in Figure 12.11. To compute the percentage correct, all students were included from countries that were included in these trend analyses. For percentage correct 25 OECD countries were included. The United Kingdom, the Netherlands, Luxembourg, the Slovak Republic and Turkey were excluded because they did not participate in either PISA 2000 or PISA 2003 or because they excluded for quality assurance reasons from either PISA 2000 or PISA 2003. In addition, recent OECD members, such as Chile, Estonia, Israel and Slovenia were not included.

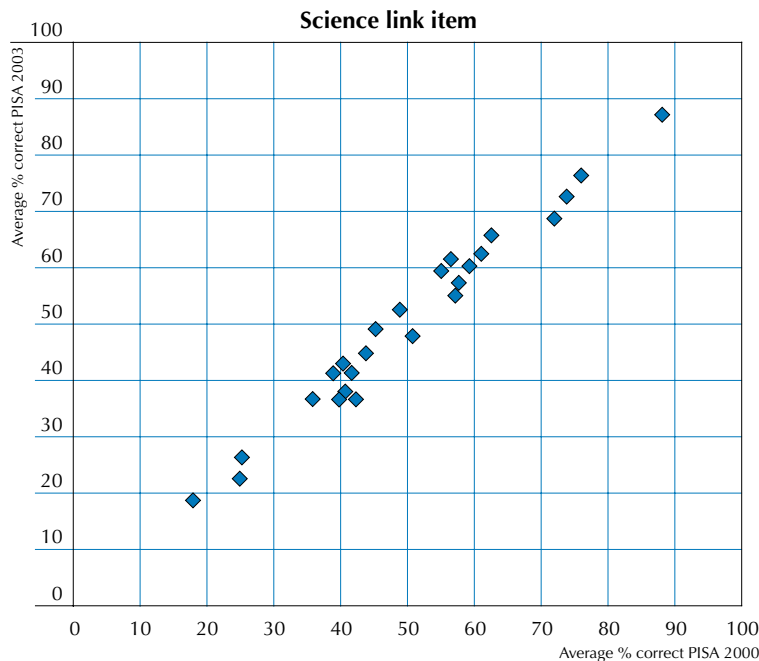
The mean of the differences (PISA 2000 minus PISA 2003) is -0.28 , and the standard deviation of the differences is 2.79. This standard deviation is consistent with that observed for reading between 2000 and 2003.

Table 12.33 International percent correct for science link items in PISA 2000 and PISA 2003

| Item | % correct | |
|---------|-----------|------|
| | 2000 | 2003 |
| S114Q03 | 57.3 | 55 |
| S114Q04 | 39.8 | 36.8 |
| S114Q05 | 24.9 | 22.7 |
| S128Q01 | 62.6 | 65.7 |
| S128Q02 | 45.2 | 49 |
| S128Q03 | 61.2 | 62.5 |
| S129Q01 | 38.8 | 41.6 |
| S129Q02 | 17.9 | 19 |
| S131Q02 | 50.9 | 47.9 |
| S131Q04 | 25.2 | 26.5 |
| S133Q01 | 56.7 | 61.6 |
| S133Q03 | 42.3 | 36.6 |
| S133Q04 | 43.8 | 44.7 |
| S213Q01 | 40.3 | 43.2 |
| S213Q02 | 76.1 | 76.6 |
| S252Q01 | 48.8 | 52.8 |
| S252Q02 | 72.2 | 68.6 |
| S252Q03 | 55 | 59.2 |
| S256Q01 | 88.3 | 87.3 |
| S268Q01 | 73.7 | 72.4 |
| S268Q02 | 40.8 | 38.1 |
| S268Q06 | 57.9 | 57.4 |
| S269Q01 | 59.2 | 60.2 |
| S269Q03 | 41.8 | 41.6 |
| S269Q04 | 35.9 | 36.5 |

■ Figure 12.11 ■

Scatter plot of percentage correct for science link items in PISA 2000 and PISA 2003



Science 2003 to 2006

In PISA 2006, science was the major domain and as such it was comprehensively assessed on the basis of a newly developed and elaborated framework. As noted above there were quite substantial changes between the preliminary framework that had underpinned PISA 2000 and PISA 2003 assessments and the more fully developed framework used for PISA 2006. Note that in addition to the framework changes mentioned above, there was an important change in the way science was assessed in PISA 2006, when compared with PISA 2003 and PISA 2000. First, to more clearly distinguish scientific literacy from reading literacy, the PISA 2006 science test items required on average less reading than the science items used in earlier PISA surveys. Second, as with each domain when it goes from a minor to a major domain the item pool, the testing experience for the majority of students becomes dominated by the new major domain. For example, there were 108 science items used in PISA 2006, compared with 35 in PISA 2003; of these, just 22 items were common to PISA 2006 and PISA 2003 and 14 were common to PISA 2006 and PISA 2000.

Therefore, as the first major assessment of science, the PISA 2006 assessment was used to establish the basis for the PISA science scale.

The percentage correct on science items that link PISA 2003 and PISA 2006 are given in Table 12.34, with the corresponding scatter plot Figure 12.12. To compute the percentage correct, all students were included from countries that were included in these trend analyses. For percentage correct, 29 OECD countries were included. The United Kingdom was excluded because it was excluded from the *PISA 2003 Database*.

The mean of the differences (PISA 2006 minus PISA 2003) is 0.01, and the standard deviation of the differences is 1.89. This standard deviation is less than for science from 2000 to 2003 but greater than that for reading from 2003 to 2006. As with the previous observations regarding the standard deviations of the differences, this is consistent with PISA test design changes.

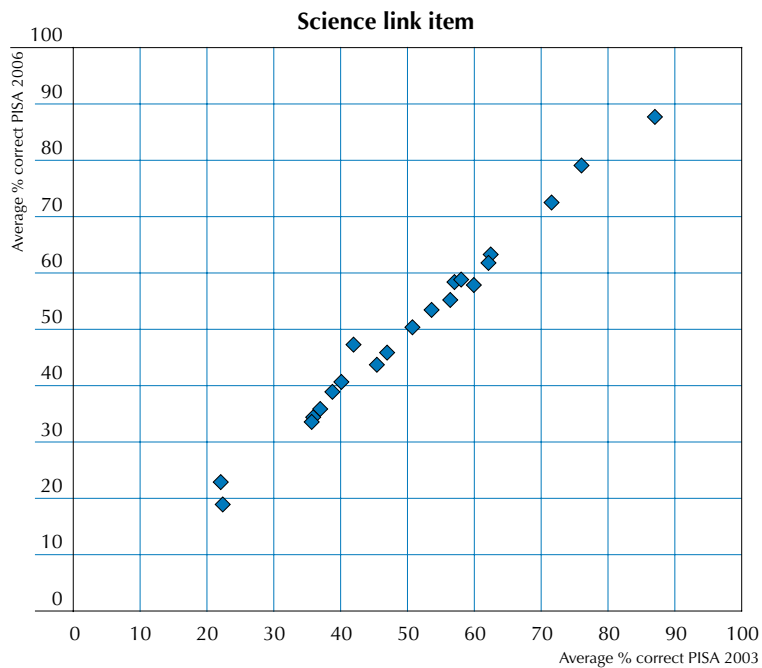


Table 12.34 International percent correct for science link items in PISA 2003 and PISA 2006

| Item | % correct | |
|----------|-----------|------|
| | 2003 | 2006 |
| S114Q03 | 53.6 | 53.6 |
| S114Q04 | 35.9 | 34.4 |
| S114Q05 | 22.4 | 18.8 |
| S131Q02 | 46.9 | 46.2 |
| S213Q01 | 41.9 | 47.4 |
| S213Q02 | 76.2 | 79.2 |
| S256Q01 | 87 | 87.5 |
| S268Q01 | 71.7 | 72.5 |
| S268Q02 | 36.9 | 36.1 |
| S268Q06 | 56.6 | 55.4 |
| S269Q01 | 60 | 57.9 |
| S269Q03 | 40.1 | 40.7 |
| S269Q04 | 35.6 | 33.8 |
| S304Q01 | 45.5 | 43.8 |
| S304Q02 | 62 | 62.1 |
| S304Q03a | 38.7 | 39.1 |
| S304Q03b | 50.7 | 50.6 |
| S326Q01 | 58.2 | 58.7 |
| S326Q02 | 62.6 | 63.4 |
| S326Q03 | 57.2 | 58.3 |
| S326Q04 | 22.2 | 22.8 |

■ Figure 12.12 ■

Scatter plot of percentage correct for science link items in PISA 2003 and PISA 2006



For the purposes of trend analysis an additional trend scale has been established that is based upon those items that were common to both PISA 2003 and 2006. Details on the construction of this trend scale are given below and international results are provided in the initial report (OECD, 2007; pp. 369-370).

On the science trend scale that was produced from these 39 countries that participated in both PISA 2003 and PISA 2006, one had a significant decline in mean score and 5 had a significant rise in mean score (OECD, 2007).

Science 2006 to 2009

Fifty-three science items were selected from PISA 2006 and used again in PISA 2009.¹³ Hence the change from 2006 to 2009 involved reducing the number of items by almost half, and as it was the case when reading and mathematics changed from major to minor domain, it was not possible to make such a reduction whilst retaining intact clusters. Three new clusters were formed for PISA 2009 from the units retained from PISA 2006.

The trend results for the OECD countries that participated in both PISA 2006 and PISA 2009 showed that of the 57 countries which could be compared 6 had a significant decline in mean score and eleven had a significant rise in mean score (OECD, 2010b).

The percentage correct on science items that link PISA 2006 and PISA 2009 are given in Table 12.35, with the corresponding scatter plot and Figure 12.13. For percentage correct, 34 OECD countries were included.

The mean of the differences (PISA 2006 minus PISA 2009) is -0.79 , and the standard deviation of the differences is 2.04.

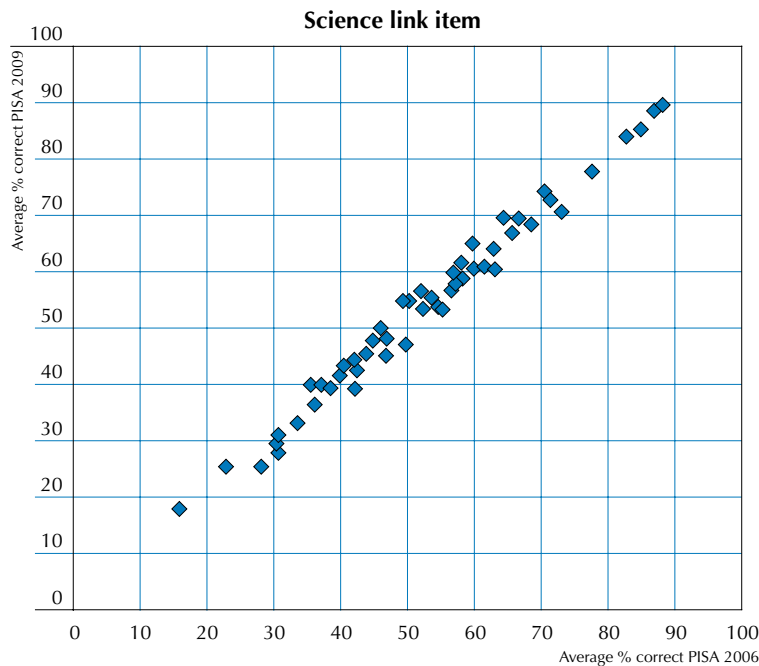
Table 12.35 International percent correct for science link items in PISA 2006 and PISA 2009

| Item | % correct | |
|----------|-----------|------|
| | 2006 | 2009 |
| S131Q02D | 46.3 | 49.5 |
| S131Q04D | 30.9 | 28.0 |
| S256Q01 | 87.4 | 88.6 |
| S269Q01 | 57.7 | 58.0 |
| S269Q03D | 40.1 | 41.4 |
| S269Q04T | 33.8 | 33.0 |
| S326Q01 | 58.3 | 58.6 |
| S326Q02 | 62.9 | 63.9 |
| S326Q03 | 57.9 | 60.6 |
| S326Q04T | 23.1 | 25.3 |
| S408Q01 | 62.8 | 60.3 |
| S408Q03 | 30.8 | 30.7 |
| S408Q04T | 50.8 | 54.4 |
| S408Q05 | 42.3 | 42.8 |
| S413Q04T | 40.8 | 43.0 |
| S413Q05 | 64.6 | 69.1 |
| S413Q06 | 37.4 | 39.7 |
| S415Q02 | 77.8 | 77.6 |
| S415Q07T | 71.4 | 72.7 |
| S415Q08T | 57.3 | 59.7 |
| S425Q02 | 45.2 | 47.4 |
| S425Q03 | 42.2 | 43.8 |
| S425Q04 | 30.6 | 29.4 |
| S425Q05 | 68.5 | 68.3 |
| S428Q01 | 61.9 | 60.5 |
| S428Q03 | 71.3 | 73.0 |
| S428Q05 | 43.7 | 45.2 |
| S438Q01T | 82.8 | 83.7 |
| S438Q02 | 65.9 | 66.7 |
| S438Q03D | 38.6 | 39.3 |
| S465Q01 | 49.8 | 46.8 |
| S465Q02 | 60.1 | 60.4 |
| S465Q04 | 36.3 | 36.2 |
| S466Q01T | 70.6 | 73.5 |
| S466Q05 | 54.8 | 53.2 |
| S466Q07T | 73.3 | 70.3 |
| S478Q01 | 42.5 | 43.0 |
| S478Q02T | 50.4 | 54.6 |
| S478Q03T | 66.9 | 69.1 |
| S498Q02T | 46.8 | 45.0 |
| S498Q03 | 42.4 | 38.9 |
| S498Q04 | 59.8 | 64.7 |
| S514Q02 | 85.0 | 84.9 |
| S514Q03 | 46.9 | 49.0 |
| S514Q04 | 52.3 | 55.9 |
| S519Q01 | 35.8 | 39.7 |
| S519Q02T | 53.6 | 54.8 |
| S519Q03 | 28.3 | 25.4 |
| S521Q02 | 54.4 | 54.2 |
| S521Q06 | 88.2 | 89.2 |
| S527Q01T | 16.1 | 17.7 |
| S527Q03T | 56.9 | 57.2 |
| S527Q04T | 52.6 | 53.1 |



■ Figure 12.13 ■

Scatter plot of percentage correct for science link items in PISA 2006 and PISA 2009



TRANSFORMING THE PLAUSIBLE VALUES TO PISA SCALES

For PISA 2009 the reading, mathematics and science results are each reported on the scales that were established when the respective domain was a major domain. Therefore in the case of reading, the results are directly comparable with those that have been reported for PISA 2000, PISA 2003 and PISA 2006. In the case of mathematics they are directly comparable with the results reported in PISA 2003 and PISA 2006 and for science they are directly comparable with the results reported in PISA 2006.

Mathematics

For mathematics, the PISA 2009 plausible values were equated to the PISA scale by using common item equating.

A shift to align the scales was computed as follows. Of the 48 mathematics items that were included in the PISA 2006 main survey, 35 were selected for PISA 2009 main survey assessment. The average item difficulty of the 35 link items was set to zero in PISA 2009 while it was 0.0752 in PISA 2006. A shift of 0.0752 of a logit was therefore required to align PISA 2006 and PISA 2009 mathematics scales. After applying this shift, the same transformation was used as in PISA 2006.

The resulting transformation required to place logits on the PISA mathematics scale was:

$$\text{PISA 2009 scaled score} = ((L + 0.1691) / 1.2838) * 100 + 500$$

where L is the logit scale outcome of the 2009 scaling.

For details about equating procedures in 2006, see the *PISA 2006 Technical Report* (OECD, 2008).

Reading

A two-step equating approach was used to report PISA 2009 reading results on the PISA 2000 reading scale.

Step 1: Common items equating

A shift to align the scales was computed as follows. Of the 101 reading items that were included in the PISA 2009 main survey, 26 were link items that had been used in each previous PISA assessment. The average item difficulty of the 26 link items was -0.0885 in PISA 2009 while in PISA 2006 it was 0.0210 . A shift of 0.0906 logits was therefore required to align the PISA 2006 and PISA 2009 reading link items.

Step 2: Common person equating

To equate PISA 2009 student proficiency scores to PISA scale, the dataset that included PISA 2009 OECD countries was scaled twice, once using all the reading items and once using only link items. The difference between the student proficiency means of these two scalings was 0.1261 logits and this shift was applied to the student PVs to place PISA 2009 student performance to the PISA scale.

After applying this shift, the transformations required to place logits on the PISA reading scale were as given below. Note that the transformation is done separately by gender, as has been the case since PISA 2003.

For female students:

$$\text{PISA 2009 scaled score} = ((0.8739 * L - 0.4416) / 1.1002) * 100 + 500$$

For male students:

$$\text{PISA 2009 scaled score} = ((0.8823 * L - 0.5185) / 1.1002) * 100 + 500$$

For students with missing gender code:

$$\text{PISA 2009 scaled score} = ((0.8830 * L - 0.4837) / 1.1002) * 100 + 500$$

Science

For science, the PISA 2009 plausible values were equated to the PISA scale by using the common items equating method.

A shift to align the scales was computed as follows. Of the 103 science items that were included in the PISA 2006 main survey, 53 were selected for the PISA 2009 main survey assessment. The average item difficulty of the 53 link items was set to zero in PISA 2009 while it was 0.0151 in PISA 2006. A shift of 0.0151 of a logit is required to align PISA 2006 and PISA 2009 science scales.

After applying this shift, the transformation required to place logits on the PISA science scales was:

$$\text{PISA 2009 scaled score} = ((L - 0.1646) / 1.0724) * 100 + 500$$

where L is the logit scale outcome of the 2009 scaling.

DRA

DRA logits were standardised to have mean of 0 and standard deviation of 1 for a combined set of 16 equal weighted OECD countries. Then the mean and standard deviation of PISA paper and pencil reading scale for this combined set was computed. Final linear transformation of the DRA logit value yields a mean and standard deviation of DRA PISA results to be equal to PISA paper and pencil reading results for a combined set.

The transformation required to place DRA logits on the PISA scales was:

$$\text{PISA 2009 scaled score} = (((L - 0.5165) / 1.1011) * 96.3956) + 498.9126$$

where L is the logit scale outcome of the 2009 scaling.

LINK ERROR

Link errors estimated using the methodology discussed in Chapter 9 were computed for the following eleven links: PISA mathematics scales 2003 to 2006, 2006 to 2009 and 2003 to 2009; PISA reading scales 2000 to 2003, 2000 to 2006, 2000 to 2009, 2003 to 2006, 2003 to 2009 and 2006 to 2009; and PISA science scale 2006 to 2009 and science trend scale 2003 to 2006. The results are given in Table 12.36.

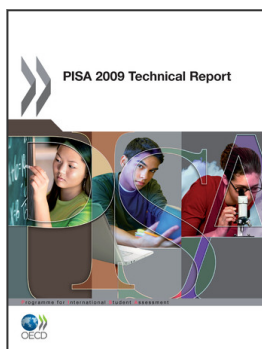
Table 12.36 Link error estimates

| | Link error on PISA scale |
|-------------------------------------|--------------------------|
| PISA mathematics scale 2003 to 2006 | 1.382 |
| PISA reading scale 2000 to 2003 | 4.474 |
| PISA reading scale 2000 to 2006 | 4.976 |
| PISA reading scale 2003 to 2006 | 5.307 |
| Interim science scale 2000 to 2003 | 3.112 |
| Science trend scale 2003 to 2006 | 4.963 |
| PISA mathematics scale 2003 to 2009 | 1.990 |
| PISA mathematics scale 2006 to 2009 | 1.333 |
| PISA reading scale 2000 to 2009 | 4.937 |
| PISA reading scale 2003 to 2009 | 4.088 |
| PISA reading scale 2006 to 2009 | 4.069 |
| PISA science scale 2006 to 2009 | 2.566 |



Notes

1. The “Xs” represent a different number of students in each graph.
2. Note that this section refers to cognitive scales only. PISA has also produced a wide range of other scales that are affective or behavioural scales.
3. This is because conditioning variables were not used in the construction of the scales for the PISA 2000 data (see *PISA 2003 Technical Report*, OECD 2005).
4. This is because gender was the only conditioning variable used in the construction of the content-based scales. (see *PISA 2006 Technical Report*, OECD 2008).
5. Note, of course, that as mentioned above comparison across alternative scalings of the same domain are not appropriate.
6. The PISA 2000 frameworks were published as OECD (1999) *Measuring Student Knowledge and Skills: A new Framework for Assessment*.
7. The PISA 2003 frameworks were published as OECD (2003) *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills*.
8. The PISA 2006 frameworks were published as OECD (2006) *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006*.
9. In 2003 the total testing time was also 390 minutes, but 60 minutes of that testing time was allocated to an assessment of problem-solving skills.
10. In 2009 the total testing time per country was also 390 minutes.
11. Representing 120 minutes of testing time.
12. Representing 90 minutes of testing time.
13. Representing 90 minutes of testing time.



From:
PISA 2009 Technical Report

Access the complete publication at:
<https://doi.org/10.1787/9789264167872-en>

Please cite this chapter as:

OECD (2012), "Scaling Outcomes", in *PISA 2009 Technical Report*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/9789264167872-13-en>

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at info@copyright.com or the Centre français d'exploitation du droit de copie (CFC) at contact@cfcopies.com.