Assessment data analysis: sample report

Social origin, school type and anxiety: France and Germany, PISA 2022

Dr Milos Kankaras | Center for Psychology

This specimen shows what a national deep dive delivers. It was not commissioned by either country. The analysis is original work on the public PISA 2022 student file.

Verdict for the minister

France (474) and Germany (475) had mathematics scores in PISA 2022 that the data cannot tell apart. Their equity gaps cannot be told apart either: the most advantaged quarter of students scored 113 points above the least advantaged in France and 111 in Germany. About half of the least advantaged did not reach Level 2 (49% in France, 47% in Germany), against 9% and 8% of the most advantaged. In the joint model, 60% of the gap in France and 64% in Germany runs between grades and school types and between schools with different social intakes. Between students of similar background in the same school type and intake, 33% and 27% remain. Placement into school types and the support given inside each type are therefore the places to look first. Do not use the published mathematics anxiety index to say that students in the two countries are equally anxious: the near-zero index difference combines questions that differ in opposite directions.

1. The equity gap

Students are divided into four equal groups by the PISA index of economic, social and cultural status (ESCS), separately in each country. Scores rise at every step in both countries (Table 1). The German quarters score 7.9 points (S.E. 4.8) above the French at the bottom and 6.8 points (S.E. 5.6) at the top; neither difference is distinguishable from zero at the 5% level. The quarters are national, so they compare students in the same relative position, not at the same absolute level of resources.

Table 1. Mathematics by socio-economic quarter (standard errors in parentheses)

	France: mean score	France: % below Level 2	Germany: mean score	Germany: % below Level 2
Least advantaged quarter	422.0 (3.0)	49.4 (1.9)	429.9 (3.8)	46.6 (2.2)
Second quarter	456.8 (3.6)	33.8 (1.8)	464.4 (4.1)	31.2 (2.0)
Third quarter	489.1 (3.4)	20.4 (1.6)	489.8 (3.9)	22.9 (1.8)
Most advantaged quarter	534.6 (3.6)	8.7 (1.2)	541.4 (4.3)	8.4 (1.3)
All students	**473.9 (2.5)**	**28.8 (1.1)**	**474.8 (3.1)**	**29.5 (1.3)**

Note. France: 6,770 students in 282 schools; Germany: 6,116 in 257. Quarters use the 6,572 and 5,314 students with an ESCS value; ESCS is missing for 3% of students in France and 13% in Germany.

The raw gap is 112.6 points in France (S.E. 4.4) and 111.5 in Germany (S.E. 5.1); the difference of 1.1 points has a standard error of 6.7, so neither country can be called less equal on this measure. The gradient tells a different story: one unit of ESCS goes with 45.5 points in France and 39.6 in Germany,

a difference of 5.9 (S.E. 2.2, z = 2.76). The two are compatible: the middle half of students spans 1.53 ESCS units in Germany and 1.34 in France, so the same quarter gap covers a wider social distance in Germany. The correct standard error matters too: a shortcut with one plausible value and no replicate weights gives 2.9 for France instead of 4.4.

Table 2. Mathematics by school type, as recorded in the PISA study programme codes

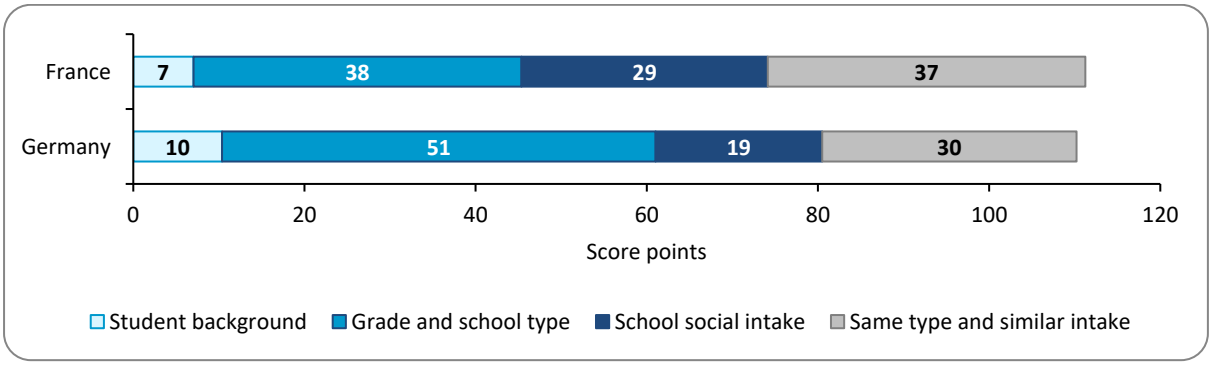
School type	Students	Share, %	Mean score (S.E.)
France: upper secondary general (ISCED 3A)	4,520	68.7	509.7 (2.2)
France: upper secondary vocational (ISCED 3B)	1,214	17.7	404.4 (3.4)
France: upper secondary vocational (ISCED 3C)	167	2.4	382.0 (6.0)
France: lower secondary (a grade or more behind)	767	9.8	391.6 (4.1)
France: lower secondary, special education	102	1.4	331.8 (7.6)
Germany: academic track (Gymnasium)	2,270	36.5	546.5 (2.6)
Germany: intermediate general track	1,551	25.0	455.5 (3.5)
Germany: comprehensive school	1,379	23.2	435.3 (4.8)
Germany: basic general track	617	10.0	396.7 (5.2)
Germany: special education	132	2.7	359.7 (14.5)
Germany: vocational, pre-vocational, Waldorf and other	167	2.6	426.7 (12.9)

Note. The codes are national. In France they describe the stage and stream of schooling, so lower secondary at this age means a delay of a grade or more. In Germany they mostly describe the track within lower secondary education.

Sorting is the visible mechanism. In France 44% of the least advantaged quarter are in general upper secondary education, against 91% of the most advantaged. In Germany 14% of the least advantaged quarter are in the academic track, against 69%. PISA has no measure of attainment before placement, so these patterns mix selection on earlier attainment with any effect of the school type itself, and they are not proven causes. The types differ in number and content between the countries and should be read within each country.

Figure 1 and Table 3 divide the raw gap among four sources, all entered in one model, so the result does not depend on the order of entry.

Figure 1. Where the gap between the most and least advantaged quarters lies, in score points



Note. The bars show the estimates in Table 3.

Table 3. Raw and adjusted gap between the top and bottom quarters, score points (S.E.)

	France	% of gap	Germany	% of gap
Raw gap, students with complete data	111.2 (4.3)	100	110.2 (4.9)	100
Gender, immigration, home language	7.0 (1.3)	6	10.4 (1.5)	9
Grade and school type	38.3 (2.7)	34	50.7 (3.8)	46
Mean ESCS of the school	28.8 (2.8)	26	19.5 (4.3)	18
Adjusted gap (remaining)	**37.1 (3.7)**	**33**	**29.7 (4.0)**	**27**

Note. Complete cases: 6,317 in France and 5,264 in Germany (93% and 86% of students). In France a grade below the modal one coincides with lower secondary, so grade enters through school type there. Adjusting in steps gives 111, 103, 53 and 37 points in France and 110, 95, 38 and 30 in Germany. Percentages are rounded and may not sum to 100.

2. Can mathematics anxiety be compared?

On the published index the two countries look the same: 0.13 in France and 0.10 in Germany, a difference of -0.03 (S.E. 0.03, z = -1.2). The test is whether the six questions work in the same way in both countries. Each student was given at most five of the six.

Table 4. Staged comparability test, France against Germany

Model and stage	CFI	RMSEA	SRMR	Change in CFI	Change in RMSEA
One factor: configural	0.967	0.244	0.105		
Two factors: configural	0.987	0.160	0.052		
Linked two factors: configural	0.999	0.051	0.014		
Same: metric (equal loadings)	0.999	0.044	0.014	0.000	-0.006
Same: scalar (equal intercepts)	0.997	0.056	0.014	-0.001	+0.012
Check with school clustering: configural	0.993	0.037	0.014		
Check with school clustering: metric	0.990	0.040	0.024	-0.003	+0.003
Check with school clustering: scalar	0.979	0.052	0.031	-0.011	+0.012

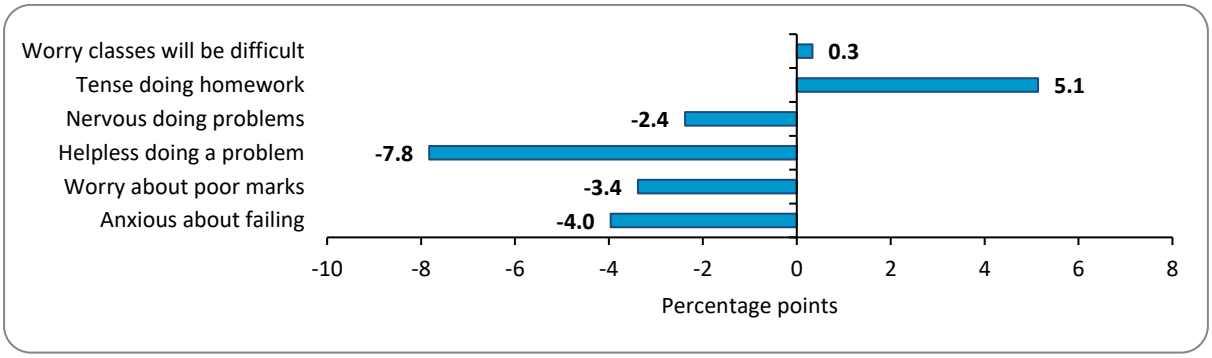
Note. Rows 1 to 5: WLSMV for ordered categories, student weights, pairwise use of the planned missing data; response thresholds are held equal before loadings. Clustered check: the same linked two-factor model, robust maximum likelihood with full information and school-clustered standard errors. n = 5,513 in France and 4,831 in Germany.

A single anxiety factor does not fit (RMSEA = 0.244). Two correlated factors, worry and tension while working, also do not fit (RMSEA = 0.160); the largest misfit is consistent with the questions on poor marks and on failing sharing more than the worry factor explains. Linking their residuals (the linked model) gives an acceptable fit (CFI = 0.999, RMSEA = 0.051). In that model equal loadings hold within the usual limits, a fall in CFI of 0.010 supplemented by a rise in RMSEA of 0.015 (Chen, 2007). Equal intercepts hold in the main test and go just past the CFI limit in the clustered check (-0.011). A report that stopped at the main test would release the country means.

The item check shows why it should not (Figure 2). Students in Germany agree 5.1 percentage points more often that they get tense doing mathematics homework and 7.8 points less often that they feel helpless doing a problem. The tension difference therefore depends on which question is trusted. With all intercepts equal, tension is -0.05 standard deviations in Germany relative to France (S.E.

0.02). If the homework question may differ it is -0.13 (S.E. 0.02); if the helplessness question may differ it is +0.02 (S.E. 0.02), not distinguishable from zero. Worry is -0.003 (S.E. 0.026) and moves between -0.046 and +0.027 when questions are released one at a time.

Figure 2. Students who agree or strongly agree, Germany minus France, in percentage points



Note. Standard errors run from 1.1 to 1.3 points.

Verdict: configural and metric, supported with reservation; scalar, not accepted. The first reservation is that the model with acceptable fit was found by changing the first one after seeing the data. The second is that fit without the link is poor (RMSEA = 0.160). Associations may be compared, with care. Country means are not released, because the clustered check rejects equal intercepts by a small margin and the tension difference changes from clearly negative to none with the choice of question. The French and German wordings were not examined, so no difference is attributed to translation; examining them is the next step.

3. Comparisons you may make and may not make

Table 5. Release list for ministry use

Statement	May state it?	Reason
The two countries' mathematics means were not distinguishable in 2022	Yes	Difference 0.9, S.E. 3.9
In each country the most advantaged quarter outscores the least advantaged quarter	Yes	France 112.6 (S.E. 4.4), Germany 111.5 (S.E. 5.1); state that quarters are national
About 60% of the French gap and 64% of the German gap lies between grades, school types and school intakes	Yes	Joint model; a description, not an effect
A unit of ESCS goes with more points in France than in Germany	With care	5.9 points, S.E. 2.2, but the quarter gap does not differ; report both
How anxiety relates to other variables, compared across the two countries	With care	Metric stage supported with reservation: model adjusted after seeing the data; change in CFI 0.000 (main) and -0.003 (clustered)
One country is less equal than the other	No	Quarter gaps differ by 1.1, S.E. 6.7; adjusted gaps by 7.4, S.E. 5.5
School type matters more for the gap in Germany than in France	No	School-type categories are national and not equivalent
Ending early tracking would close 51 points of the German gap	No	One survey of 15-year-olds without prior attainment cannot show this
Students in the two countries are equally anxious about mathematics	No	Equal index values hide opposite question differences: +5.1 and -7.8 points

4. Methods and data

Data. PISA 2022 student questionnaire file, public use (OECD, 2023a), all sampled students in France and Germany. Immigrant background is missing for 5% in France and 12% in Germany, home language for 2% and 11%. Grade and study programme are complete, and all six anxiety questions are available in both countries. School mean ESCS uses the sampled students.

Scores and errors. Each statistic is computed 810 times: for each of ten plausible values, with the final student weight and with each of 80 replicate weights. Sampling variance uses balanced repeated replication with Fay's factor of 0.5 and is averaged over plausible values; imputation variance is added by Rubin's rules. Country samples are independent, so variances add in differences. Quarter boundaries are recomputed with every replicate weight, also in the decomposition; a boundary student goes to the lower quarter. ESCS and school means are fixed.

Validation. The code reproduces the published PISA 2022 mathematics means and standard errors to two decimal places: France 473.94 (2.49) and Germany 474.83 (3.06) (OECD, 2026, Table I.B1.2a.38). It also reproduces the share of students at Level 1a, 17.85% (0.68) and 17.97% (0.84) (OECD, 2023b, Table I.B1.3.1), and the share of score variance explained by ESCS, 21.48% (1.31) and 18.70% (1.28) (Table I.B1.5.22). Quarter gaps and their standard errors match the published values within 0.01: France 112.58 (4.36), Germany 111.50 (5.08) and Austria, used only as a test of the code, 106.49 (3.98) (Table I.B1.5.19). Random order among students with equal ESCS (five seeds) moves no gap by more than 0.06 and no gap S.E. by more than 0.01.

Comparability test. lavaan 0.6-21 and semTools 0.5-8 in R 4.5.3. Standard errors of the latent differences in the text come from refitting the two-factor model without the link with all 80 replicate weights; with the link, tension estimates differ by less than 0.001. Without the link, the clustered check gives changes in CFI of -0.007 and -0.013. The Chen (2007) cutoffs were developed with maximum likelihood on continuous, normally distributed data and are a guide here.

What the numbers support. They describe 15-year-olds in school in 2022, the size and location of the equity gap, and the safe use of one questionnaire index. They do not identify policy effects.

Dr Milos Kankaras, Founder and Director, Center for Psychology | info@centerforpsychology.me

Specimen: a client report carries the reviewer's dated signature here.

References

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464-504. https://doi.org/10.1080/10705510701301834

OECD. (2023a). *PISA 2022 database* [Data set]. https://www.oecd.org/en/data/datasets/pisa-2022-database.html

OECD. (2023b). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

OECD. (2026). *PISA 2025 results (Volume I): Future-ready students*. OECD Publishing. https://doi.org/10.1787/73451bc5-en