Economics Bulletin

Volume 44, Issue 4

How reliable are financial-literacy country rankings?

Leo Van Hove
Vrije Universiteit Brussel (VUB)

Muzaffarjon Ahunov
Woosong University

Abstract

Gignac and Ooi (2022) show that financial literacy tests composed of fewer than ten questions suffer from low internal consistency reliability. This note shows that the issue also limits the reliability and relevance of country rankings based on such tests.

We would like to thank Allianz, and in particular Lorenz Weimann, for allowing us to use the data, and Petra Brandes, Michela Coppola, Andrea Hasler, and Annamaria Lusardi for helping us get hold of them. Thanks also go to Kiril Kossev for facilitating access to the OECD data and to Marc Jegers for his comments.
Citation: Leo Van Hove and Muzaffarjon Ahunov, (2024) "How reliable are financial-literacy country rankings?", *Economics Bulletin*, Volume 44, Issue 4, pages 1457-1463
Contact: Leo Van Hove - Leo.Van.Hove@vub.be, Muzaffarjon Ahunov - mahunov@wsu.ac.kr.
Submitted: August 26, 2024. **Published:** December 30, 2024.

How reliable are financial-literacy country rankings?

1. Introduction: the (potential) problem

To assess individuals' financial literacy, the academic literature relies heavily on simple metrics such as the Big Three, Standard & Poor's, and OECD/INFE. These survey instruments consist of, respectively, three, five, and seven questions on basic financial concepts such as compound interest, inflation, and risk diversification [1].

Importantly, the use of the instruments is not limited to national surveys; they are also used in multi-country efforts. The resulting 'league tables' understandably attract considerable attention, including in the international press. For example, on the publication of the 2014 Standard & Poor's Global Financial Literacy Survey – which covers no less than 143 countries – *The Wall Street Journal* headlined: "The U.S. May Be the World's Richest Country, But It Ranks 14th in Financial Literacy" [2].

However, a recent article has a potentially inconvenient message. Gignac and Ooi (2022) compute the internal consistency reliability – the Cronbach's α or a comparable measure – of 52 samples of financial literacy estimates, for either individual countries or a pool of countries. (In the latter case, they aggregate the scores of individual respondents across countries.) The financial literacy estimates rely on a variety of tests, including more comprehensive instruments than the ones already mentioned.

Gignac and Ooi (2022) find that studies that make use of the Big Three have reliabilities between 0.30 and 0.47 – "values less than minimally acceptable for even exploratory research" (o.c., p. 938). Overall, many estimates based on fewer than 10 test items – including the 7-item OECD/INFE questionnaire – have reliability values below the 0.70 recommended for early-stage research. Gignac and Ooi (o.c., p. 948) conclude that "researchers should […] avoid using a financial literacy test with only three items, or even five items, despite the popularity of the Big Three and Big Five" [3]; they advise to use minimum 13-15 questions.

The single-country evaluations of Gignac and Ooi (2022) also raise questions as to the reliability of country rankings. However, *a priori* it is not clear what to expect: if the internal consistency reliabilities are similarly low across countries, then the countries' relative positions might not be affected much, if at all. We therefore perform explicit tests. We find – and this is our contribution compared to Gignac and Ooi (2022) – that the low internal consistency of simple financial literacy scales at the country level can indeed also invalidate the often-used country rankings, even though there are exceptions.

2. Data

In what follows, we test, in a cross-country setting, two of the simple metrics criticised by Gignac and Ooi (2022), namely the Big Three and the OECD/INFE Toolkit. We would have liked to also exploit the 2014 Standard & Poor's survey – which has the broadest geographical coverage – but the micro data are not publicly available.

---

[1] For a comparison, see Van Hove and Ahunov (2024).

[2] Source: *WSJ*, November 18, 2015. The Standard & Poor's results can be found in Klapper, Lusardi, and Van Oudheusden (2015).

[3] The 'Big Five' questionnaire supplements the Big Three with questions on bond prices and mortgages; see Global Financial Literacy Excellence Center, "The Big Three and Big Five", at url: https://gflec.org/education/questions-that-indicate-financial-literacy/ (accessed October 18, 2024).

For our test of the Big Three, we make use of a survey conducted by financial services company Allianz (2017) at the same point in time – in November 2016 – in 10 western European countries. The samples are nationally representative with respect to age, sex, and geography. The FLat World project also employs the Big Three and covers more countries – 15 to be precise – but consists of separate surveys conducted by different teams of researchers between 2007 and 2014 (Lusardi, 2019). Also, in some of the surveys the wording of the questions was slightly different from the original.

The OECD/INFE Toolkit, for its part, has been applied to a large number of countries but this was done in multiple waves between 2015 and 2023 (Van Hove and Ahunov, 2024). It therefore did not seem advisable to aggregate all the results in one big ranking. Moreover, the availability of the micro data is patchy. We obtained access to the results of the survey conducted in 2021 in the Commonwealth of Independent States (CIS) (OECD, 2021). This dataset consists of only 8 countries but has the advantage that the countries are at a different stage of economic development compared to those in the first dataset.

Table 1 presents the country ranking that stems from the Allianz survey. We base the ranking on the average scores in column (4), rather than on the values in column (8) that are typically reported in league tables, namely the percentage of the population that is financially literate. In the Big Three test, an individual needs to answer all three questions correctly for them to be considered literate. As the use of such a threshold entails a transformation of the data, we prefer to work with the raw scores [4]. However, our findings also hold for the numbers in column (8).

Table 1. Allianz financial literacy estimates

The table reports detailed results for the 10 countries covered in the Allianz (2017) survey, which uses the Big Three measure. N = sample size; Average score = number of correct answers out of 3; % literate = percentage of respondents who answered all three questions correctly. Scores and confidence intervals calculated with sampling weights.

Rank	Country	N	Average score	Standard deviation	95% confidence interval		% literate
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	AT	1,008	2.15	0.96	2.10	2.21	47
2	DE	1,032	2.15	0.94	2.09	2.21	46
3	CH	1,008	2.06	0.94	2.00	2.11	40
4	ES	1,072	1.89	0.95	1.84	1.95	31
5	NL	1,021	1.84	1.02	1.78	1.90	33
6	BE	1,015	1.82	0.98	1.76	1.88	29
7	UK	1,041	1.77	0.94	1.71	1.82	24
8	IT	1,015	1.76	0.95	1.70	1.82	25
9	FR	1,018	1.74	1.01	1.68	1.81	28
10	PT	1,050	1.72	1.03	1.66	1.79	28

In much the same way as Table 1, Table 2 presents the OECD (2021) results. In the OECD/INFE approach, the threshold to be considered literate is five out of seven correct answers [5]. Sample size is 1,000 per country (o.c., p. 18).

[4] See Ahunov and Van Hove (2024) for a detailed justification. Note that the rankings in columns (4) and (8) differ. The Spearman correlation coefficient is 0.87. By contrast, because of the higher variance in the scores, in Table 2 the rankings in columns (3) and (7) are identical.

[5] To be clear: the full OECD/INFE Toolkit contains questions on not only financial knowledge, but also on financial attitudes and behaviors. However, for comparability, we only use the (seven) questions on financial knowledge.

Table 2. OECD financial literacy estimates

The table reports detailed results for the 8 countries covered in the OECD (2021) survey, which uses the OECD/INFE measure. Average score = number of correct answers out of 7; % literate = percentage of respondents who answered five out of the seven questions correctly. All statistics are calculated with sampling weights.

Rank	Country	Average score	Standard deviation	95% confidence interval		% literate
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	Belarus	4.5	1.63	4.40	4.61	50
2	Kazakhstan	4.2	1.70	4.10	4.31	44
3	Russian Federation	4.1	1.64	4.01	4.21	42
4	Uzbekistan	3.6	1.55	3.50	3.69	31
5	Armenia	3.5	1.49	3.40	3.59	26
6	Kyrgyz Republic	3.2	1.52	3.15	3.34	22
7	Tajikistan	3.0	1.49	2.91	3.09	16
8	Azerbaijan	2.4	1.80	2.34	2.56	14

3. The tests

With the above data, we perform a test inspired by Gignac and Ooi (2022), the goal being to ascertain the confidence with which the country rankings of the previous section may be interpreted. We first report, in detail, the results obtained with the Allianz data and subsequently, in brief (and more as a robustness check), the results for the OECD/INFE metric.

Concretely, for each country in Table 1 we compute, first, the internal consistency reliability of the test scores (as measured by the coefficient alpha) and, second, the standard error of measurement (SEM):

$$\text{SEM} = \text{SD}\ \sqrt{1-\alpha}, \qquad (1)$$

where SD = the standard deviation of the scores, and α = the Cronbach's alpha of the sample.

As can be seen in column (3) of Table 3, the Cronbach's alphas – which, as Gignac and Ooi (2022, p. 951) stress, are sample-specific (and not innate to any given test) – range between 0.34 and 0.54. This is slightly higher than what Gignac and Ooi (2022) find for their Big Three samples (see Section 1), but still largely insufficient.

Column (4) shows that the standard errors of measurement are large compared to the observed scores (which range between 0 and 3). As a result, the 95% confidence intervals in columns (5) and (6) are wide – to say the least. To take an extreme case, the true score for Italy would lie somewhere between 0.25 and 3.27. As a matter of fact, the upper bound of the confidence interval exceeds the maximum value of 3 for all countries in the table.

Figure 1 shows that the confidence intervals are much wider than those in Table 1, which do not take into account the Cronbach's alphas. This underpins Gignac and Ooi's (2022, p. 952) call for researchers "to report statistical results that take into consideration the measurement error of their test scores".

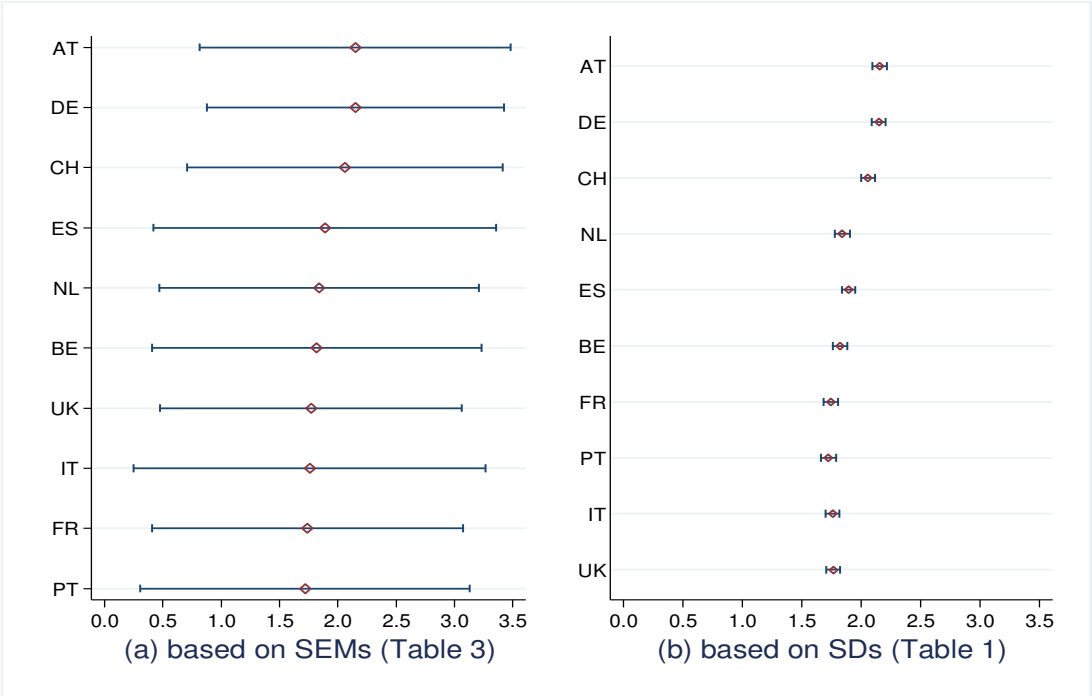
Table 3. Allianz estimates: reliability and confidence intervals

The table presents reliability measures and confidence intervals for the Allianz financial literacy estimates reported in Table 1.

Country	Average score	α	SEM	95% confidence interval	
(1)	(2)	(3)	(4)	(5)	(6)
AT	2.15	0.50	0.68	0.82	3.49
DE	2.15	0.52	0.65	0.87	3.42
CH	2.06	0.47	0.69	0.71	3.40
ES	1.89	0.39	0.75	0.43	3.36
NL	1.84	0.54	0.70	0.48	3.21
BE	1.82	0.46	0.72	0.41	3.23
UK	1.77	0.51	0.66	0.46	3.07
IT	1.76	0.34	0.77	0.25	3.27
FR	1.74	0.54	0.68	0.41	3.08
PT	1.72	0.51	0.72	0.31	3.14

Figure 1. Allianz estimates: 95% confidence intervals

The left-hand panel of the figure depicts confidence intervals for the Allianz estimates based on SEMs, as reported in columns (5) and (6) of Table 3. The right-hand panel depicts confidence intervals based on SDs, as reported in columns (6) and (7) of Table 1.



The above obviously does not bode well for the reliability of the country ranking in Table 1. Table 4 presents the original ranking alongside rankings based on, respectively, the lower and upper bounds of the confidence intervals listed in Table 3. In other words, in column (3) of Table 4 we assume that, for all countries, the true score coincides with the lower bound of the confidence interval. In column (4), we assume that it lies at the higher bound. The idea is to gauge just how much the country ranking might change if one takes into account the margin of error of the test scores, as highlighted by Gignac and Ooi (2022). Let us stress that while the two scenarios might appear to be extreme cases, it is, in fact, a relatively lenient test. Indeed,

it is not a given that the measurement errors go in the same direction for all countries. That is, the true score of one country might be overestimated, and that of the next underestimated. As can be gleaned from the left-hand panel of Figure 1, such a scenario could shake up the ranking even more.

But even in our lenient test, the original ranking is impacted substantially: countries can move up to three places in either direction; compare column (2) of Table 4 with columns (3) and (4). The Spearman correlation coefficients in the bottom row provide a quantification of the differences. In interpreting these coefficients, one should keep in mind the leniency of our test. Note that the Spearman correlation between the upper- and lower-bound rankings is only 0.59.

Note also that a similar test with the confidence intervals in Table 1 yields Spearman coefficients of 1, as the country ranking is unaffected. This again highlights the critical impact of the Cronbach's alphas.

Table 4. Allianz country rankings: original vs. sensitivity analysis
The table compares the original country ranking based on the Allianz (2017) report, in column (2), with rankings that take into account the margin of error in the test scores, in columns (3) and (4). Spearman coefficients are correlations with the original ranking.

		Rank based on	
Country	Average score	Lower bound	Upper bound
(1)	(2)	(3)	(4)
AT	1	2	1
DE	2	1	2
CH	3	3	3
ES	4	6	4
NL	5	4	7
BE	6	7	6
UK	7	5	10
IT	8	10	5
FR	9	8	9
PT	10	9	8
Spearman		0.88	0.84

To return to Table 3 (and Figure 1), with confidence intervals that overlap to such an extent, perhaps the conclusion should be even harsher. Indeed, one could question whether, with the Big Three data that we have, it makes any sense to construct country rankings based on point estimates, as several of the inter-country differences may well be insignificant.

As announced, we perform a similar test for the OECD/INFE metric with the data in Table 2. Table 5 summarises the results. In line with Gignac and Ooi (2022, p. 946), the Cronbach's alphas are higher than for the Big Three. Still, they are always lower than 0.70. The standard errors of measurement in column (4) are larger compared to those in Table 3, but here one has to take into account that the financial literacy scores now range between 0 and 7.

Table 5. OECD estimates: reliability and confidence intervals
The table presents reliability measures and confidence intervals for the OECD financial literacy estimates reported in Table 2.

Country	Average score	α	SEM	95% confidence interval	
(1)	(2)	(3)	(4)	(5)	(6)
Belarus	4.5	0.55	1.10	2.36	6.65
Kazakhstan	4.2	0.64	1.03	2.19	6.22
Russian Federation	4.1	0.57	1.08	2.00	6.22
Uzbekistan	3.6	0.51	1.08	1.47	5.71
Armenia	3.5	0.49	1.07	1.40	5.59
Kyrgyz Republic	3.2	0.53	1.04	1.20	5.29
Tajikistan	3.0	0.46	1.09	0.87	5.13
Azerbaijan	2.4	0.69	1.01	0.48	4.42

Most importantly, as can be gleaned from inspection of columns (5) and (6), if one were to rank the countries based on the lower and upper bounds of the 95% confidence intervals – in the same way as in Table 4 – the rankings would, with one small exception, be identical to the original. (The exception is that in the ranking based on the upper bounds, the Russian Federation would – narrowly – jump from place 3 to place 2.) As a result, the Spearman correlation coefficients are, respectively, 1.00 and 0.97.

Crucially, however, the explanation for the fact that, unlike for the countries in the Allianz dataset, the ranking of the CIS countries remains largely unaffected does not lie in the higher reliability of the financial literacy estimates. As mentioned, all Cronbach's alphas are below the 0.70 that is considered to be satisfactory. Also, there is again considerable variation across the samples. Rather, the explanation is that the spread in the original financial literacy estimates is substantially higher compared to the Allianz dataset; compare columns (2) in Tables 5 and3.

Finally, let us warn against drawing too much comfort from the near-perfect Spearman correlation coefficients. As stressed earlier, our test is relatively lenient and the confidence intervals in columns (5) and (6) are wide.

4. Discussion and conclusion

We are not the first to criticise the popular financial-literacy country rankings. In particular, de Clercq (2019) performs psychometric tests on the seven financial knowledge questions of the OECD/INFE Toolkit. Her analysis, with data for 11 out of the 30 countries that participated in the 2015 wave, shows that item difficulties are not homogeneous across the various countries. de Clercq concludes that the metric does not qualify as an 'International Large-Scale Assessment' because comparability across countries is not guaranteed.

The present note reveals an additional reason to treat country rankings based on simple tests with caution: the internal consistency reliabilities of the country scores are not only low, they may also vary to such an extent that the true ranking can be substantially different. In other words, because the financial literacy scales typically have low internal consistency on the country level, rankings that, by definition, build on country-level estimates are not reliable. A possible exception is when the distances between the financial literacy scores of the individual countries are large to start with.

To wrap up, as Grohmann, Klühs, and Menkhoff (2018, p. 86) point out (concerning the Standard & Poor's metric), financial literacy scales and their items are often "simplified,

probably [with an eye on] wider coverage of countries beyond advanced economies". However, the present note shows that comparability is not guaranteed. Needless to say, this finding is potentially problematic for much of the cross-country research on financial literacy, including our own [6].

References

Ahunov, M. and L. Van Hove (2020) "National culture and financial literacy: international evidence" *Applied Economics* **52**, 2261-2279.

Ahunov, M. and L. Van Hove (2024) "Big Three illiterate, Standard & Poor's literate? Experimental evidence on differences between financial literacy metrics" Mimeo.

Allianz (2017) "When Will the Penny Drop? Money, Financial Literacy and Risk in the Digital Age – 2017 Report" International Pension Papers 1/2017

De Clercq, B. (2019) "A comparative analysis of the OECD/INFE financial knowledge assessment using the Rasch model" *Empirical Research in Vocational Education and Training* **11**, 1-29.

Gignac, G. E. and E. Ooi (2022) "Measurement error in research on financial literacy: How much error is there and how does it influence effect size estimates?" *Journal of Consumer Affairs* **56**, 938-956.

Grohmann, A., Klühs, T. and L. Menkhoff (2018) "Does financial literacy improve financial inclusion? Cross country evidence" *World Development* **111**, 84-96.

Klapper, L., Lusardi, A. and P. Van Oudheusden (2015) "Financial Literacy Around the World: Insights from Standard & Poor's Ratings Services Global Financial Literacy Survey" Global Financial Literacy Excellence Centre Working Paper

Klapper, L. and A. Lusardi (2020) "Financial literacy and financial resilience: evidence from around the world" *Financial Management* **49**, 589-614.

Lusardi, A. (2019) "Financial literacy and the need for financial education: evidence and implications" Swiss Journal of Economics and Statistics **155**, 1-8.

Meoli, M., Rossi, A. and S. Vismara (2022) "Financial literacy and security-based crowdfunding" *Corporate Governance* **30**, 27-54.

OECD (2021) "Financial Literacy Levels in the Commonwealth of Independent States in 2021" www.oecd.org/financial/education/financial-literacy-levels-in-the-commonwealth-of-independant-states-2021.htm

Van Hove, L. and M. Ahunov (2024) "Financial literacy: different indicator, different insights?" *Cogent Economics & Finance* **12**, 2437005.

[6] Besides Ahunov and Van Hove (2020), examples include Grohmann, Klühs, and Menkhoff (2018), Klapper and Lusardi (2020), and Meoli, Rossi, and Vismara (2022) – all of which rely on Standard & Poor's financial literacy estimates.