

# Are You Financially Literate or Not? Experimental Evidence on Discrepancies between Two Metrics

Muzaffarjon AHUNOV - Endicott College of International Studies, Woosong University, Jayang-Dong, Dong-gu Daejeon, South Korea (mahunov@endicott.ac.kr)  
*corresponding author*

Leo VAN HOVE - Department of Applied Economics (APEC), Vrije Universiteit Brussel (Free University of Brussels), Brussels, Belgium

## *Abstract*

*This note reports on an online, within-subjects experiment with a convenience sample of 206 respondents. The experiment consisted in having the same set of respondents taking not one but two of the financial literacy tests that are popular in the literature, namely the Big Three and the Standard & Poor's. Disturbingly, we find that 37 per cent of the respondents are considered literate by one test but illiterate by the other. One explanation is that the difficulty level of questions that are relatively similar across the two approaches would nevertheless appear to differ. Another explanation involves the minimum number of correct answers needed for someone to be classified as financially literate. We show that, at least for our sample, a solution might consist in removing these thresholds. The differences in raw scores between the two tests proved to be not statistically significant.*

## **1. Introduction**

In a recent paper, Van Hove and Ahunov (2024) replicate two cross-country studies on financial literacy, and replace the Standard & Poor's indicator used in the original studies by, alternately, the Big Three and OECD/INFE metrics. Disturbingly, they find that several results change depending on how financial literacy is measured. As their replications make use of overlapping samples, Van Hove and Ahunov conclude that the differences in results must necessarily be caused by differences in the country-level financial literacy estimates; in other words, by the measurement tools.

This note reports on a small-scale, exploratory experiment that we set up to shed light on where the differences between the metrics might come from. The experiment consisted in having the same set of respondents taking not one but two financial literacy tests. Concretely, our online survey contained both the S&P as well

---

<https://doi.org/10.32065/CJEF.2025.01.03>

Conflict/Declaration of Interest Statement: To the best of our knowledge we have no conflict of interest. We declare that we had no external funding for this research.

We would specifically like to thank the two anonymous referees and the handling editor for their valuable comments and recommendations.

as the Big Three questions. To the best of our knowledge, this is the first time the financial literacy of individuals is measured by means of two different metrics simultaneously.

We find that the S&P approach produces a markedly higher literacy rate compared to the Big Three: 82 vs. 55 per cent. This discrepancy arises because 32 per cent of our respondents are classified as financially literate under the S&P test, but as illiterate under the Big Three. Conversely, another 5 per cent is deemed literate by the Big Three but not by the S&P approach. As such, these inconsistent classifications ‘contribute’ to the literature that questions the reliability of simple financial literacy tests (de Clercq, 2019; Nicolini and Haupt, 2019; Gignac and Ooi, 2022).

On a technical level, there are two main explanations for the differences in literacy rates we observe. First, the difficulty level of questions that are (more or less) similar across the two approaches would nevertheless appear to differ. This may be due to the use of specific terminology. Second, there is the difference in ‘thresholds’. The threshold is the minimum number of questions an individual needs to answer correctly to be considered financially literate (three out of the four topics in the S&P approach vs. all three questions in the Big Three). We show that, while adjusting the thresholds predictably affects the literacy rates, it does not fully eliminate the differences between the approaches. Moreover, the choice of threshold is inherently arbitrary, as there is no *a priori* criterion to determine the most appropriate value. A novel finding is that removing the thresholds altogether, and working with ‘raw’ scores, would seem to be a better, albeit not perfect option.

In what follows, Section 2 first provides details on the two financial literacy metrics that are compared in our experiment. For completeness, the OECD/INFE indicator is also briefly discussed. Subsequently, Section 3 sets out the inspiration for the experiment and Section 4 describes the set-up. Finally, Section 5 reports and analyses the results, and Section 6 concludes.

## 2. S&P, Big Three, and OECD/INFE

The metric of financial literacy (in fact: financial knowledge) that is most popular in the literature is the so-called Big Three. The Big Three is a set of three questions initially developed in 2004 by Lusardi and Mitchell (Lusardi, 2019, p. 2). The questions test individuals’ knowledge of compound interest, inflation, and risk diversification. As Lusardi (o.c.) explains, between 2009 and 2014 the Big Three questions have been added to national surveys in 15 countries, as part of the FLat World project <sup>2</sup>. Together with two more recent, international surveys – by Allianz and Aegon – this gives Van Hove and Ahunov (2024) a dataset of 26 unique countries, albeit with measurements at different points in time.

For its part, the 2014 Standard & Poor’s Global Financial Literacy Survey (Klapper, Lusardi, and Van Oudheusden, 2015) was a most welcome resource for the then still young literature on financial literacy, as it removed the bottleneck for cross-country analysis presented by the limited availability of data. The survey was administered by S&P Ratings Services in cooperation with Gallup, Inc., the

---

<sup>2</sup> In some of the FLat World surveys, the questions did have slightly different wording compared to the original Big Three.

World Bank, and the Global Financial Literacy Excellence Centre (GFLEC). It probed 150,000 adults, across 148 countries – the result being, in the words of GFLEC, the ‘largest, most global measurement of financial literacy’.

The S&P survey looks at the same three concepts as the Big Three, but has two questions on compound interest instead of one, and also tests respondents’ basic numeracy. According to Grohmann, Klühs, and Menkhoff (2018, p. 86), “[w]hile Klapper et al. (2015) do not explicitly discuss the exact origins of their [S&P] survey questions, it is quite obvious that the three questions on risk diversification, inflation and interest compounding are slight variations of the standard items used in the literature”; that is, the Big Three. The differences in the wording of the overlapping questions can be ascertained in the first two columns of Table A1 in the Appendix. Note that the S&P measure should not be confused with the ‘Big Five’ developed by Lusardi and Mitchell. The Big Five test consists of the Big Three (in the exact same wording) supplemented with a question on bond prices and one on mortgages<sup>3</sup>.

Even though we do not use the metric in our experiment, let us – for completeness – also mention the Survey of Adult Financial Literacy Competencies of the OECD/INFE (International Network on Financial Education). This is an OECD-coordinated international effort to assess financial literacy in a standardised way, by means of the OECD/INFE ‘Toolkit’ – a survey blueprint. The OECD/INFE defines financial literacy as a “complex phenomenon, made up of a combination of knowledge, attitudes and behaviours” (OECD, 2016, p. 52). The Toolkit therefore contains questions on all three dimensions. For reasons of comparability, Van Hove and Ahunov, in their replication efforts, only use the (seven) questions on financial knowledge (OECD, 2016, p. 20). Three of these – on compound interest, inflation, and risk diversification – are comparable (but not identical) to the Big Three; see again Table A1. However, the OECD/INFE Toolkit has a second item on inflation, there is an item on the concept of interest, as in the S&P survey there is a numeracy question, and, finally, there is an item on the relationship between risk and return that has no equivalent in either the Big Three or S&P measures.

Initially, 38 countries participated in the OECD survey – in two waves. Later, the toolkit was also used in regional surveys in Eurasia and South East Europe, as well as in a third and fourth global wave; see Van Hove and Ahunov (2024) for details.

### 3. Inspiration

As mentioned in the Introduction, the present note was inspired by the fact that Van Hove and Ahunov (2024) find different results when they substitute the S&P financial literacy metric used in the studies they replicate by, alternatively, the Big Three and OECD/INFE indicators. Along the road, Van Hove and Ahunov (o.c., p. 5) observe that “the cross-country correlations between the S&P, Big Three and OECD/INFE financial literacy estimates are far from perfect. In other words, it is not merely a matter of one measure consistently yielding higher estimates than the other; the relative positions of the countries differ too”. This raises the question of where

---

<sup>3</sup> Global Financial Literacy Excellence Center (GFLEC), “The Big Three and Big Five”, at url: <https://gflec.org/education/questions-that-indicate-financial-literacy/> (last accessed on 3.12.2024).

these differences might come from, especially since, as pointed out in the previous Section, the tests are actually quite similar.

In this respect, Van Hove and Ahunov observe that, if they focus on the sample of 19 countries for which they have estimates for all three metrics, the Big Three estimates are almost invariably lower than the S&P estimates. In turn, the S&P estimates are in most cases (14 out of the 19, to be precise) lower than the OECD/INFE scores.

Van Hove and Ahunov point out that, at first sight, this would seem to be in line with the thresholds that are used in the different tests. Obviously, *ceteris paribus*, the higher the threshold, the lower the probability that a person is classified as literate by that measure. The respective thresholds are: all three questions in the Big Three approach, 3 out of the 4 topics for the S&P measure, and 5 out of the 7 questions for the OECD/INFE metric. The required ‘scale saturation’ would thus seem to be, respectively, 100, 75, and 71 per cent. However, both the S&P and the OECD/INFE measure have additional stipulations. As a result, the real minimum levels are, in fact, 100 per cent, 60-80 per cent <sup>4</sup>, and 71-86 per cent <sup>5</sup>. The upshot is that while the difference in thresholds helps explain why country-level Big Three estimates tend to be lower than those based on the S&P test, it cannot explain why S&P estimates are usually lower than OECD/INFE ones.

Also, the different measures do purport to be able to discriminate between financially literate and illiterate individuals. One would thus assume that the threshold that was picked takes into account the nature and difficulty of the questions. Indeed, concerning the S&P measure, Klapper, Lusardi, and Van Oudheusden (2015, p. 7) state the following: “A person is defined as financially literate when he or she correctly answers at least three out of the four financial concepts [...]. We choose this definition because the concepts are basic and this is what would correspond to a passing grade”. Also regarding the S&P measure, Klapper and Lusardi (2020, p. 592) explain: “Because interest compounding is such a hard concept, [...] respondents need to answer only one out of two questions correctly to get a score of one” <sup>6</sup>. In any case, one should not forget the ‘*ceteris paribus*’ condition. The differences in estimates between the measures could also be due to differences in the nature of the survey questions.

#### 4. The Experiment

To explore the above further, we have run a small-scale online experiment. In the course of February 2020 we sent the members of the professional social network of the lead author a link to a questionnaire. The questionnaire contained both the five

---

<sup>4</sup> In the S&P test, the compound interest topic consists of two questions, and a respondent is considered as having understood this concept as soon as they correctly answer one (Klapper and Lusardi, 2020, p. 592). In other words, three out of five correct answers (60 per cent) can be enough to be classified as financially literate. But if a respondent answers the two interest questions correctly, they need another two correct answers – or four in total (80 per cent).

<sup>5</sup> Here one has to take into account that the OECD considers the answer to the compound interest question correct only if the respondent also correctly answers the question on simple interest.

<sup>6</sup> In a recent article, Demertzis, Mejino-López, and Moffat (2024, p. 4) opt for the three-out-of-five threshold simply because this “is consistent with other work measuring financial knowledge across countries”. They do so even though it is not a given that the difficulty level of the five questions in the European Commission survey that they analyse is similar to that of the five S&P questions.

S&P questions as well as the Big Three. To the best of our knowledge, this is the first instance the financial literacy of a set of individuals is measured by means of two different metrics at the same time <sup>7</sup>. (Testing three metrics in one experiment did not seem viable, so we opted for the two most similar.)

To be clear: our experiment is not a randomised controlled trial (RCT), where the sample is randomly split in two (or more) groups. Rather we opted for a within-subjects experiment, where all respondents are exposed to every condition. We did so because, given our sampling procedure, we anticipated a relatively low number of respondents. For the composition of the groups in an RCT to be balanced (in terms of socio-demographic characteristics, etc.), a sufficiently high number of subjects is needed, as the law of large numbers needs to take effect.

Conversely, a drawback of a within-subjects experiment lies in the potential order effects. Order effects occur when the sequence in which the conditions are presented – in our case, say, first the Big Three and then the S&P test – affects participants’ performance. Subjects may learn with more exposure, or they might become bored. To control for this, we resorted to (complete) ‘counterbalancing’; that is, we randomised the order, not just of the two sets of questions, but of all the individual questions in the two sets. Most respondents will thus have received the Big Three and S&P questions intermingled. We did so to ensure that any order-related effects are distributed across the conditions, thus reducing their potential impact on the test results. In addition, we complemented the financial literacy questions with other questions on household financial behaviour <sup>8</sup>, the idea being to make the purpose of the survey less obvious. The order of these questions was also randomised. As a result, they will typically have shown up in between the real test questions.

The online nature of the survey is another potential drawback of our set-up. As Kalmi and Ruuskanen (2018, p. 338) note, face-to-face interviews “are often regarded as the ‘gold standard’ in financial literacy studies” because other survey formats allow respondents to look up information or answers. On the upside, online surveys do not suffer from the interviewer effects documented by Crossley et al. (2021).

Finally, let us mention that even though the majority of the potential respondents were not native speakers of English, the questionnaire was solely in English. But given the nature of the population (mostly academics, professionals with a European or US degree, and students following economics and business programmes where English is the medium of instruction) we did not anticipate this to be an issue. In any case, the within-subjects nature of our experiment should also limit the problem, as any variability between individuals – including abilities – is controlled for.

In total we received 260 responses, 206 of which were complete and usable. Table A2 contains descriptive statistics. Due to the procedure followed, our sample is

---

<sup>7</sup> As documented by Van Hove and Ahunov (2024), there are authors who do use more than one measure. However, their approach consists in extending an existing metric with additional questions, rather than, say, employing the S&P measure as an alternative to the Big Three.

<sup>8</sup> The additional questions were taken from the World Values Survey. The entire questionnaire is available upon request.

evidently not representative of any population. Men and highly-educated individuals are overrepresented, and the vast majority have a bank account. Conversely, there are no elderly people in our sample. The lack of representativeness limits the external validity of our results but, crucially, does not compromise their internal validity. Indeed, for our purposes, the sample does not need to be representative. Unlike in the surveys discussed in Section 2, our goal was not to come up with a reliable estimate of the overall level of financial literacy in any country. Rather we wanted to test whether, for a given group of people, the use of a different set of questions might result in a different financial literacy score – and, if possible, find out why.

## 5. Results

### 5.1. Test Scores

Table 1 shows the percentage of correct answers for each of the questions in the S&P and Big Three tests, as well as, in the bottom rows, the overall financial literacy rates (in bold). As an aside, note that, in line with the observations of Gignac and Ooi (2022), the Cronbach alphas of the scales are low, substantially below the recommended 0.7.

**Table 1 Financial Literacy Experiment: % of Correct Answers and Overall Scores**

Questions	S&P				Big Three				Difference in correct answers (a) – (e)
	Correct	Incorrect	Do not know	Refuse to answer	Correct	Incorrect	Do not know	Refuse to answer	
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	
Risk diversification	82	13	4	1	64	14	19	3	18 <sup>***</sup>
Inflation	66	29	4	1	79	13	7	2	-13 <sup>**</sup>
Numeracy	82	12	4	2					
Compound interest, Q1	76	16	7	1					
Compound interest, Q2	78	19	2	1	93	4	1	1	-16 <sup>***</sup>
Cronbach's alpha	0.49				0.47				
Financial literacy	<b>82</b>				<b>55</b>				27 <sup>***</sup>

Source: own survey. See Table A1 for wording of the questions. Percentages may not sum up to 100 due to rounding. <sup>\*\*\*</sup> p < 0.001, <sup>\*\*</sup> p < 0.01. The survey sample consists of 206 respondents with an undergraduate degree or higher. Descriptive statistics can be found in Table A2.

To focus, for now, on the overall test scores, let us stress that the S&P approach yields a markedly higher literacy rate than the Big Three: 82 vs. 55 per cent. In view of Van Hove and Ahunov's observation, mentioned in Section 3, that the country-level Big Three estimates are almost consistently lower than the S&P estimates, this was to be expected.

Table 2 cross-tabulates the individual-level 'verdicts' of the two tests; that is, whether or not an individual is considered financially literate. As can be seen, no less than 37 per cent of the respondents (32.04 + 5.34) are deemed literate by one test, but

illiterate by the other. Given that the tests purport to measure the same construct, this is troubling.

**Table 2 Cross-Tabulation of Results of Big Three and S&P Tests: Pearson  $\chi^2(1) = 12.63$ ; Pr = 0.00**

		S&P			
		<i>Illiterate</i>	<i>Literate</i>	<i>Total</i>	
<i>Big Three</i>	<i>Illiterate</i>	%	<b>13.11</b>	<b>32.04</b>	<b>45.15</b>
		N	(27)	(66)	(93)
	<i>Literate</i>	%	<b>5.34</b>	<b>49.51</b>	<b>54.85</b>
		N	(11)	(102)	(113)
	<i>Total</i>	%	<b>18.45</b>	<b>81.55</b>	<b>100</b>
		N	(38)	(168)	(206)

Notes: Numbers in bold are percentages; number of observations in parentheses.

The observations in Tables 1 and 2 are obviously interrelated: it is precisely because the S&P test classifies significantly more individuals as financially literate (Table 1) that there are so many cases where the two tests disagree (Table 2). For the same reason, the bulk of the cases relate to individuals who are deemed literate in the S&P approach but illiterate by the Big Three (33 per cent of all respondents; again Table 2). To be clear: this should not be taken as implying that the S&P overestimates financial literacy; it is also possible that the Big Three underestimates it. Note in this respect that there are also individuals, albeit fewer (some 5 per cent of all respondents), where the discrepancy between the two tests goes in the opposite direction. In other words, it is not solely a situation of one test ostensibly being more lenient than the other.

The above results raise the questions of why the S&P and Big Three approaches yield different financial literacy rates, and why so many respondents are classified inconsistently. Let us therefore delve deeper into Table 1 and examine the scores per question.

## 5.2. Difficulty of Individual Questions

Perhaps somewhat surprisingly, it can be seen that the higher overall result for the S&P test does not really come from higher shares of correct answers on the three questions that are comparable with the Big Three; see also Tables A3-A5. In fact, the ‘easiest’ question – with 93 per cent correct answers – proves to be the Big Three question on compound interest<sup>9</sup>. The Big Three inflation question is also correctly answered by a significantly higher share of respondents than the corresponding S&P question (79 vs. 66 per cent). Conversely, the score on the risk diversification question is significantly lower for the Big Three variant. Interestingly, here 19 per

<sup>9</sup> This could be because \$102 – see Table A1 – is not particularly well chosen as a benchmark. As was also pointed out to us by a number of respondents, \$110 would have made it a better test of respondents’ understanding of compound interest. Now it would seem more of a numeracy test. Tellingly, in the recent pan-European Union survey carried out by the European Commission, which uses €110 as the benchmark, the compound interest question is the least well understood of the three ‘Big Three’-like questions (Demertzis et al., 2024, pp. 16-17).

cent of the respondents answered “Do not know”, which might indicate that the question lacks clarity. The problem might be the presence of the terms “stock” and “mutual fund”, since the S&P variant of the question contains no such terminology; see Table A1. As Grohmann, Klühs, and Menkhoff (2018, p. 86) point out in their comments on the origin of the S&P questions, “The item on risk diversification has been simplified [compared to the Big Three], probably to reflect the wider coverage of countries beyond advanced economies”.

Broadening the scope to the S&P questions that have no counterpart in the Big Three still does not explain the higher overall literacy rate: across the five S&P items, the percentage of correct answers is 77 per cent, compared with 79 per cent for the Big Three. Also, the success rates for the S&P numeracy question (82 per cent) and the first compound interest question (76 per cent) do not diverge much from the Big Three average. This suggests that the difference in threshold between the S&P and Big Three approaches might matter.

### 5.3. Impact of the Thresholds

As explained in detail in Section 3, the threshold for an individual to be classified as literate is lower in the S&P than in the Big Three: three out of five correct answers can be enough. In the Big Three, the respondent needs a perfect score. In the counterfactual exercise in Table 3, we therefore treated the answers to the S&P questions *à la Big Three*. That is, we focused on the three S&P questions that match (more or less) with the Big Three, and applied the higher Big Three threshold – thus raising the required scale saturation from 60-80 per cent to 100 per cent (see Section 3).

Unsurprisingly, this lowers the S&P estimate, from 82 per cent literate individuals to 46 per cent. The impact is straightforward to understand. With five questions, there are 32 possible ‘answer patterns’ in the S&P test; that is, sequences of correct and incorrect answers. Crucially, in the original S&P approach all five questions matter, but in our ‘amended’ S&P test only the answers to the three that are similar to the Big Three count. With this in mind, it is intuitive that there are literally no cases where an individual would be considered literate by the amended S&P but not by the full S&P: having the three ‘Big Three’-like questions correct suffices to meet the three-out-of-four-topics S&P threshold. A simple simulation (available upon request) shows that for 23 of the 32 answer patterns (or 72%) the verdict of the two tests is the same. In the remaining 28%, the individual is considered literate by the full S&P but not by the amended version. These are all cases where the simulated test person has one or more of the non-‘Big Three’ questions correct, but gives a wrong answer on at least one of the ‘Big Three’-like questions (where there is no room for mistakes). The crux is that the amended S&P test never gives a better result than the full test, gives the same result for 72 per cent of the answer combinations, and a worse result for 28 per cent.

This explains why it is only natural that treating the S&P answers *à la Big Three* results in fewer respondents being classified as literate. However, strikingly, our counterfactual exercise not only lowers said proportion, but lowers it significantly *below the Big Three estimate* (of 55 per cent). This implies that more of our respondents make at least one mistake on the three questions taken from the S&P test than on the Big Three questions. This is another indication that some of the



S&P questions are more difficult than their Big Three counterpart, at least in our sample; see the final column of Table 1.

**Table 3 Cross-Tabulation of Results of Big Three and S&P Tests, with S&P Test Made as Similar as Possible to Big Three: Pearson  $\chi^2(1) = 18.83$ ; Pr = 0.00**

		S&P redux – all three Big Three questions correct			
		Illiterate	Literate	Total	
Big Three	Illiterate	%	<b>32.04</b>	<b>13.11</b>	<b>45.15</b>
		N	(66)	(27)	(93)
	Literate	%	<b>22.33</b>	<b>32.52</b>	<b>54.85</b>
		N	(46)	(67)	(113)
	Total	%	<b>54.37</b>	<b>45.63</b>	<b>100</b>
		N	(112)	(94)	(206)

Notes: Numbers in bold are percentages; number of observations in parentheses.

Because of this, the problem of the inconsistent classification of respondents is still there in Table 3: now 35 per cent (13.11 + 22.33) would be classified differently by the two tests, which is only slightly lower than the 37 per cent in Table 2. A (predictable) difference compared to Table 2 is that it is now the S&P test that categorises relatively more respondents as financially illiterate (as the threshold has been raised).

In Table 4a, we have tried yet more alternatives for the S&P threshold. In Table 4b, we have also lowered the Big Three threshold. The underlying idea was that the higher S&P literacy rate compared to the Big Three (82 per cent vs. 55 per cent) could stem from the S&P threshold being too low, the Big Three threshold being too high, or a combination of both.

Horizontally, Tables 4a and 4b are divided in two panels. In the top panel, we use all five S&P questions; in the bottom panel, we work only with the three ‘Big Three’-like questions. In column (a), the thresholds are arranged from the strictest to the most lenient. The threshold in bold is the original threshold.

**Table 4a Simulations with Different Thresholds for S&P; Big Three: 3 Out of 3 (54.58 % Literate)**

Threshold	S&P % Literate	Difference with Big Three (b) - 54.58	% Classified inconsistently		
			S&P literate, Big Three illiterate (d)	Big Three literate, S&P illiterate (e)	Sum (d) + (e) (f)
(a)	(b)	(c)	(d)	(e)	(f)
All 5 questions	35.92	-18.93 <sup>***</sup>	7.28	26.21	33.49
All 4 concepts	44.66	-10.19 <sup>*</sup>	13.11	23.20	36.31
4 out of 5 questions	66.99	12.14 <sup>**</sup>	21.84	9.71	31.55
<b>3 out of 4 concepts<sup>a</sup></b>	<b>81.55</b>	<b>26.70<sup>***</sup></b>	<b>32.04</b>	<b>5.34</b>	<b>37.38</b>
All Big Three questions <sup>b</sup>	45.63	-9.22 <sup>*</sup>	13.11	22.33	35.44
2 out of 3 Big Three	83.98	29.13 <sup>***</sup>	32.52	3.40	35.92

Notes: <sup>a</sup> See also Table 2. <sup>b</sup> See also Table 3. Bold indicates that this is the threshold in the original test.  
<sup>\*\*\*</sup> p < 0.001, <sup>\*\*</sup> p < 0.01, <sup>\*</sup> p < 0.05.

**Table 4b Simulations with Different Thresholds for S&P; Big Three: 2 out of 3 (83.50 % Literate)**

S&P		% Classified inconsistently			
Threshold	% Literate	Difference with Big Three (b) - 54.58	S&P literate, Big Three illiterate	Big Three literate, S&P illiterate	Sum (d) + (e)
(a)	(b)	(c)	(d)	(e)	(f)
All 5 questions	35.92	-47.58 <sup>***</sup>	0.49	48.06	48.55
All 4 concepts	44.66	-38.83 <sup>***</sup>	2.43	41.26	43.69
4 out of 5 questions	66.99	-16.50 <sup>***</sup>	4.37	20.87	25.24
<b>3 out of 4 concepts</b>	81.55	-1.94	8.74	10.68	19.42
All Big Three questions	45.63	-37.86 <sup>***</sup>	2.43	40.29	42.72
2 out of 3 Big Three	83.98	0.49	8.74	8.25	16.99

Notes: Bold indicates that this is the threshold in the original test. \*\*\* p < 0.001, \*\* p < 0.01, \* p < 0.05.

In both tables and all panels, the pattern in the results is clear. When the S&P threshold is lowered, % Literate, in column (b), increases and the difference with the Big Three, in column (c), narrows – at least at first. In Table 4a, at a certain point, as the S&P threshold becomes even more lenient, the discrepancy starts increasing again – but now in the opposite direction.

The discrepancy between the two tests is lowest when the Big Three threshold is lowered to “2 out of 3 questions” (in Table 4b), combined with a S&P threshold of “3 out of 4 concepts” (for the test with five questions) or “2 out of 3 questions” (for the test with 3 questions). In other words, the discrepancy is minimised when there is some leniency in the thresholds for both tests. However, there is no way to verify whether the corresponding literacy rates – 81.55 to 83.98 per cent – are accurate. Note also that while the discrepancy in overall literacy rate between the two tests may be insignificant, the proportion of respondents classified inconsistently remains quite high; see column (e).

#### 5.4. What About Raw Scores?

In Table 5, we have recalculated the test results a final time, now based on the approach proposed by Henry, Huynh, and Welte (2018, p. 3). Henry et al. calculate two scores: *score 1* is simply the number of correct answers; *score 2* is the sum of correct answers minus the sum of incorrect answers, not counting Don’t know (and, in our case, also Refuse to answer) responses. The idea behind *score 2*, which will obviously be lower than *score 1*, is to penalise respondents who guess or are overconfident (Henry et al., 2018, p. 3, footnote 2)<sup>10</sup>.

Given the difference in the number of questions (five vs. three), the raw scores for the original S&P test cannot be compared *as is* with those for the Big Three. We have therefore added percentage scores (in the final column). The key observation is that the *score 1* results are not significantly different. This is true not only when all five S&P questions are taken into account (76.6 per cent vs. 78.5 for the Big Three),

<sup>10</sup> Cwynar, Cwynar, and Wais (2019), in their study on debt literacy in Poland, use the same approach as Henry et al. for their *score 2*. Cwynar et al. also have a variant where they subtract two points per incorrect answer rather than one – so as to penalise guessing even more. Note, however, that Cwynar et al.’s measure of debt literacy consists of twelve questions.

but also for the three-item S&P test (2.25 on 3 vs. 2.35; 75.1 per cent vs. 78.5). For the ‘reduced’ S&P test even the standard deviation is similar to that for the Big Three.

**Table 5 Financial Literacy Scores Following the Approach of Henry et al. (2018)**

	Mean	Standard deviation	Min	Max	% score	Difference (p value)		
						S&P-5	S&P-3	Big Three
<b>S&amp;P – Five questions</b>								
Score 1	3.83	1.21	0	5	76.6	-	1.52 (0.12)	-1.88 (0.30)
Score 2	3.06	1.79	0	5	61.2		1.07 (0.48)	-9.77*** (0.00)
<b>S&amp;P – Big Three questions</b>								
Score 1	2.25	0.83	0	3	75.1		-	-3.40 (0.10)
Score 2	1.81	1.17	0	3	60.3			-10.84*** (0.00)
<b>Big Three</b>								
Score 1	2.35	0.82	0	3	78.5			-
Score 2	2.13	1.11	0	3	71.0			

Notes: Score 1 = total number of correct answers. Score 2 = sum of correct answers minus sum of incorrect answers; no points for ‘Don’t know’ and ‘Refuse to answer’ responses; score set to zero for respondents with more incorrect than correct answers. *p* values are for t-tests. \*\*\* *p* < 0.001.

These observations suggest that when comparing S&P and Big Three test results, it is perhaps best not to use any threshold, and simply use the raw scores. A caveat is that – not dissimilar to the inconsistent classification of individuals when using thresholds (Tables 4a and 4b) – the correlation between the S&P and Big Three individual-level results for *score 1* is far from perfect. The correlation between the raw scores on the original S&P test and on the Big Three is 0.50 (*p* < 0.01); between the three-item S&P test and the Big Three it is 0.43 (*p* < 0.01). On the positive side, these correlations are higher than those between the S&P and Big Three literacy rates obtained with the ‘optimal’ thresholds identified in Table 4b. In both cases, the correlation is 0.38<sup>11</sup>.

On a final note, for *score 2* the S&P results are significantly below that for the Big Three. This can be explained by the higher proportion of incorrect answers for the S&P questions on compound interest and especially inflation; see Table 1.

## 6. Conclusions

Inspired by the findings of Van Hove and Ahunov (2024), for the present note we conducted a small-scale, exploratory experiment to investigate the sources of differences between two popular financial literacy metrics, namely the S&P and the Big Three.

Given that we rely on a convenience sample, the results of our experiment are not conclusive, but they do highlight a number of avenues for future research. For one, the experiment would seem to show that the difficulty level of the questions that

<sup>11</sup> Chi-squared tests produced consistent results.

are (more or less) similar across the S&P and Big Three approaches can nevertheless differ significantly – perhaps because of the use of specific terminology.

Another issue that is worthy of further research is the selection of the threshold level of correctly answered questions. We show that, unsurprisingly, the choice of threshold has a substantial impact on the literacy rate. More interestingly, we also show that specific combinations of thresholds can remove the discrepancy in literacy rates between the two tests. However, the internal validity of the resulting literacy rates cannot be ascertained. Perhaps most interestingly, our analysis questions whether thresholds should be used at all and suggests that raw scores might offer a better alternative. Put differently, our analysis suggests that a continuous classification may be preferable to a binary approach.

Our experiment could be build upon and improved in several ways. For one, it would be interesting to conduct a similar analysis involving the OECD/INFE metric. Question framing, which we did not touch upon, is another interesting dimension. The most obvious extension would be to repeat our analysis in a full-scale survey, with a representative sample – either as a within- or a between-subjects experiment. Such an analysis might also yield useful insights for curriculum development.

But most importantly, despite its limitations, just like Van Hove and Ahunov (2024) – and in line with earlier critical voices discussed therein – this note too sends a strong signal that the short and simple financial literacy measurement tools that have almost become informal standards in the literature might simply be too simple.

## APPENDIX

Table A1 lists and compares the questions used in the S&P, Big Three and OECD/INFE surveys, so as to highlight similarities and differences. Table A2 presents descriptive statistics for our on-line experiment; see Section 4 in the main text. Finally, Tables A3-A5 crosstabulate the answers to individual Big Three and S&P questions.

**Table A1 Measuring Financial Literacy: S&P vs. Big Three vs. OECD/INFE**

S&P	Big Three	OECD/INFE
How many can answer 3 out of these 4 topics correctly?	How many can answer all 3 questions correctly?	Number of correct answers out of 7 Words or phrases in < > can be edited to fit the national context. Italics = alternative wording available to simplify the language where necessary.
<b>RISK DIVERSIFICATION</b>		
Suppose you have some money. Is it safer to put your money into one business or investment, or to put your money into multiple businesses or investments? [one business or investment; <b>multiple businesses or investments</b> ; don't know; refused to answer]	Please tell me whether this statement is true or false. "Buying a single company's stock usually provides a safer return than a stock mutual fund." [true; <b>false</b> ; do not know; refuse to answer]	Is the following statement true or false? It is usually possible to reduce the risk of investing in the stock market by buying a wide range of stocks and shares or <i>It is less likely that you will lose all of your money if you save it in more than one place.</i> [true; false; do not know; refuse to answer]
<b>INFLATION</b>		
Suppose over the next 10 years the prices of the things you buy double. If your income also doubles, will you be able to buy less than you can buy today, the same as you can buy today, or more than you can buy today? [less; <b>the same</b> ; more; don't know; refused to answer]	Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, how much would you be able to buy with the money in this account? [more than today; exactly the same; <b>less than today</b> ; do not know; refuse to answer]	Imagine that five <brothers> are given a gift of <\$>1,000 and have to share the money equally. The <brothers> have to wait for one year to get their share of the <\$>1,000 and inflation stays at <x>%. In one year's time, will they be able to buy [Multiple choice; correct response depends on inflation used]
		Is the following statement true or false? High inflation means that the cost of living is increasing rapidly. [true; false; do not know; refuse to answer]

**Table A1 Measuring Financial Literacy: S&P vs. Big Three vs. OECD/INFE (continued)**

S&P	Big Three	OECD/INFE
<p>NUMERACY (INTEREST)</p> <p>Suppose you need to borrow 100 US dollars. Which is the lower amount to pay back: 105 US dollars or 100 US dollars plus three percent? [105 US dollars; <b>100 US dollars plus three percent</b>; don't know; refused to answer]</p>		
<p>INTEREST</p> <p>You lend \$25 to a friend one evening and he gives you \$25 back the next day. How much interest has he paid on this loan? [Open response; correct response: <b>'none'/0</b>]</p>		
<p>COMPOUND INTEREST</p> <p>Suppose you put money in the bank for two years and the bank agrees to add 15 percent per year to your account. Will the bank add more money to your account the second year than it did the first year, or will it add the same amount of money both years? [<b>more</b>; the same; don't know; refused to answer]</p>		
<p>Suppose you had 100 US dollars in a savings account and the bank adds 10 percent per year to the account. How much money would you have in the account after five years if you did not remove any money from the account? [<b>more than 150 dollars</b>; exactly 150 dollars; less than 150 dollars; don't know; refused to answer]</p>	<p>Suppose you had \$100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow? [<b>more than \$102</b>; exactly \$102; less than \$102; do not know; refuse to answer]</p>	<p>(continuation of numeracy (interest) question) ... and how much would be in the account at the end of five years [add if necessary: remembering there are no fees or tax deductions]? Would it be ... [Multiple choice; correct response: <b>more than \$110</b>, but only taken into account if simple interest question is correct]</p>
<p>RISK AND RETURN</p> <p>An investment with a high return is likely to be high risk or <i>If someone offers you the chance to make a lot of money it is likely that there is also a chance that you will lose a lot of money.</i> [<b>true</b>; false; do not know; refuse to answer]</p>		
Source: Klapper, Lusardi, and Van Oudheusden (2015, p. 6)	Source: Lusardi (2019, p. 2, Table 1)	Source: OECD (2016, p. 20, Table 1)

Notes: Answer options are in the brackets, with the correct answer in bold. Questions are not placed under headings in OECD (2016); headings have been added.

**Table A2 S&P vs. Big Three Experiment: Descriptive Statistics (N = 206)**

	<i>Frequency/ (Mean)</i>	<i>Standard Deviation</i>	<i>Min</i>	<i>Max</i>
(Age)	31.68	8.48	19.00	59.00
Female	0.38	0.49	0.00	1.00
Married	0.46	0.50	0.00	1.00
Undergraduate degree	0.36	0.48	0.00	1.00
Postgrad	0.64	0.48	0.00	1.00
Has a bank account	0.93	0.26	0.00	1.00
Migrant	0.32	0.47	0.00	1.00
<b>Country of current residence</b>				
Uzbekistan	0.29	0.45	0.00	1.00
Albania	0.15	0.36	0.00	1.00
South Korea	0.15	0.36	0.00	1.00
Turkey	0.08	0.27	0.00	1.00
United States	0.06	0.24	0.00	1.00
Pakistan	0.05	0.22	0.00	1.00
Other countries	0.22	0.22	0.00	1.00

Source: Own survey. d

**Table A3 Cross-Tabulation of Answers to the Big Three and S&P Risk Diversification Questions: Pearson  $\chi^2(1) = 14.40$ ; Pr = 0.00**

		<b>S&amp;P</b>			
		<i>Incorrect</i>	<i>Correct</i>	<i>Total</i>	
<i>Big Three</i>	<i>Incorrect</i>	%	<b>11.65</b>	<b>24.76</b>	<b>36.41</b>
		N	(24)	(51)	(75)
	<i>Correct</i>	%	<b>6.8</b>	<b>56.8</b>	<b>63.59</b>
		N	(14)	(117)	(131)
	<i>Total</i>	%	<b>18.45</b>	<b>81.55</b>	<b>100</b>
		N	(38)	(168)	(206)

Note: Numbers in bold are percentages; number of observations in parentheses.

**Table A4. Cross-Tabulation of Answers to the Big Three and S&P Inflation Questions: Pearson  $\chi^2(1) = 4.71$ ; Pr = 0.03**

		<b>S&amp;P</b>			
		<i>Incorrect</i>	<i>Correct</i>	<i>Total</i>	
<i>Big Three</i>	<i>Incorrect</i>	%	<b>10.19</b>	<b>11.17</b>	<b>21.36</b>
		N	(21)	(23)	(44)
	<i>Correct</i>	%	<b>23.79</b>	<b>54.85</b>	<b>78.64</b>
		N	(49)	(113)	(162)
	<i>Total</i>	%	<b>33.98</b>	<b>66.02</b>	<b>100</b>
		N	(70)	(136)	(206)

Note: Numbers in bold are percentages; number of observations in parentheses.

**Table A5 Cross-Tabulation of Answers to the Big Three and S&P Compound Interest Questions: Pearson  $\chi^2(1) = 10.50$ ; Pr = 0.00**

			<b>S&amp;P</b>		
			<b>Incorrect</b>	<b>Correct</b>	<b>Total</b>
<i>Big Three</i>	<i>Incorrect</i>	%	<b>3.88</b>	<b>2.91</b>	<b>6.8</b>
		N	(8)	(6)	(14)
	<i>Correct</i>	%	<b>18.45</b>	<b>74.76</b>	<b>93.2</b>
		N	(36)	(154)	(192)
	<i>Total</i>	%	<b>22.33</b>	<b>77.67</b>	<b>100</b>
		N	(46)	(160)	(206)

Notes: Numbers in bold are percentages; number of observations in parentheses. For S&P, we have taken the second compound interest question; see Table A1.



## REFERENCES

- Crossley TF, Schmidt T, Tzamourani P, Winter JK (2021): Interviewer Effects and Measurement of Financial Literacy. *Journal of the Royal Statistical Society: Series A (Statistics and Society)*, 184(1):150-178.
- Cwynar A, Cwynar W, Wais K (2019): Debt Literacy and Debt Literacy Self-Assessment: The Case of Poland. *Journal of Consumer Affairs*, 53(1):24-57.
- 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*, 11:8.
- Demertzis M, Mejino-López J, Moffat LL (2024): The State of Financial Knowledge in the European Union: A New Survey. *Journal of Financial Literacy and Wellbeing*, 2(1):38-62.
- Gignac GE., Ooi E (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(2):938-956.
- Grohmann A, Klühs T, Menkhoff L (2018): Does Financial Literacy Improve Financial Inclusion? Cross Country Evidence. *World Development*, 111:84-96.
- Henry CS, Huynh KP, Welte A (2018): 2017 Methods-Of-Payment Survey Report, Staff Discussion Paper 2018-17, Bank of Canada.
- Kalmi P, Ruuskanen OP (2018): Financial Literacy and Retirement Planning in Finland. *Journal of Pension Economics & Finance*, 17(3):335-362.
- Klapper L, Lusardi A, Van Oudheusden P (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, Lusardi A (2020): Financial Literacy and Financial Resilience: Evidence from Around the World, *Financial Management*, 49(3):589-614.
- Lusardi A (2019): Financial Literacy and the Need for Financial Education: Evidence and Implications. *Swiss Journal of Economics and Statistics*, 155:1.
- Nicolini G, Haupt M (2019): The Assessment of Financial Literacy: New Evidence from Europe. *International Journal of Financial Studies*, 7(3):54.
- OECD (2016): OECD/INFE International Survey of Adult Financial Literacy Competencies, OECD, Paris, [www.oecd.org/finance/OECD-INFE-International-Survey-of-Adult-Financial-Literacy-Competencies.pdf](http://www.oecd.org/finance/OECD-INFE-International-Survey-of-Adult-Financial-Literacy-Competencies.pdf)
- Van Hove L, Ahunov M (2024): Financial Literacy: Different Indicator, Different Insights?, *Cogent Economics & Finance*, 12(1):2437005.