ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0076

Cross-country Comparability and Validity of Metacognition of Assessing Online Information Credibility in PISA 2018: Evidence from 37 OCED Countries

Hangyan Yu[1, a]

[1]Department of Linguistics, School of International Studies, Zhejiang University, Hangzhou City, Zhejiang Province, 310058, China

[a]yhyan@zju.edu.cn

Abstract

To address the challenges posed by the rich inventories of knowledge created by complex or even contradictory information online, the metacognition of assessing online information credibility is supremely important. To inform researchers the components or prerequisites of this metacognition and how it is related to other learning aspects so that educators could construct effective pedagogical intervention to facilitate students' learning, accurate measurement of students' metacognition is important. As an influential educational benchmark for policy makers and educators worldwide, Programme for International Student Assessment (PISA) explicitly administered a new item measuring students' metacognition of assessing online information credibility, the validity of which have not been explicitly addressed Therefore, this study aimed to testify the validity of this newly-developed item on the metacognition of assessing online information credibility based on the theoretical construct among metacognition, cognition and self-efficacy. Multilevel mediation analysis of 294,527 samples from 37 Organization for Economic co-operation and Development (OECD) countries proved that the item of metacognition of assessing online information credibility in PISA is valid across different countries. This finding could facilitate future studies on this particular metacognition and some peripheral studies on how it is related to other learning parameters so that more insightful understanding on the metacogniton of assessing online information credibility could be achieved. The cross-national comparability of the validity of metacognition of assessing information credibility could also facilitate educators and policy makers worldwide to conduct effective pedagogical measures on students' digital literacy cultivation.

Keywords

Cognition; Digital literacy; Metacognition; PISA; Self-efficacy.

1. Introduction

With the rapid development of information and communications technology (ICT), there is a notable shift of reading from printed to digital texts. Digital reading, as defined in the PISA context, requires readers to be minimally ICT competent, to search for texts online, to assess information quality and credibility, to corroborate information, and to resolve potential discrepancies and conflicts [1]. It has been noted that the additional complexities of hypertexts compared to printed texts—a wealth of information, smaller displays, cluttered screens and networks of pages, required additional metacognition [2]. To address the challenges posed by the rich inventories of knowledge created by complex or even contradictory information online, it is essential for learners to metacognitively supervise the process of assessing online information credibility [3, 4]. Therefore, the role of metacognition in assessing online

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0076

information credibility in students' digital reading performance is particularly significant as a topic for further investigation in the digital era.

To inform researchers the components or prerequisites of the metacognition and how it is related to other learning aspects so that educators could construct effective pedagogical intervention to facilitate students' learning, accurate measurement of students' metacognition is important [5]. In testing, the measurement of students' targeted abilities involves two distinct aspects of identifying, estimating, and interpreting sources of variance in test scores, namely, reliability and validity [6]. The estimation of test reliability quantifies the variance that is attribute to the factors other than measurement errors such as some undesirable variances caused by test method facets or random factors. In contrast, validity measures how much the abilities that researchers aim to test contribute to the reliable variance [6]. In the measurement of students' metacognition, a cross-national assessment named Programme for International Student Assessment (PISA) explicitly administered items to test students' metacognitive strategy use in reading since 2009. In the previous rounds of PISA, two kinds of metacognition have been measured, namely, (1) the metacognition of understanding and remembering, and (2) the metacognition of summarizing. The reliability and validity of these two metacognitions have been tested in the previous studies and the results were in the acceptable range [5, 7]. The item for evaluating the metacognition of assessing credibility was used for the first time in the PISA 2018, the validity of which remains to be examined.

Considering the importance of metacognition of assessing online information credibility for students' digital reading performance, and the significance of testifying the validity of newly-developed items for tests, especially for large-scale test such as PISA, which has a relatively strong impact on educational systems worldwide [8], the current research aimed to explore the cross-country comparability and validity of metacognition of assessing online information credibility.

2. Literature Review

Validity has been traditionally classified into different types, in which content, criterion, and construct validity are the major concerns of researchers [6]. Content validity is concerned with the content relevance and content coverage of test tasks, in which the former requires test tasks should be related to the specification of the ability domain, and the latter demands test tasks should adequately represent the domain in question [6]. Criterion validity testifies the relationship between test scores and some criterion which is also an indicator of the ability tested, for example, students' performance in the tasks that involve this ability. Construct validity investigates the extent to which performance on tests is consistent with predictions that have been made based on the theory of abilities, or constructs [6].

A considerable literature has focused on the investigation of testifying the validity of PISA items. A wide range of items have been covered in PISA validity testing, such as collaborative problem solving [9], gender issues [10], opportunity to learn [11] etc. In addition, attention has been paid to the validity of different task type of PISA, such as the use of open-ended items [12]; the validity of different scaling methods, for example, the comparison of validity of PISA items using anchored scales and non-anchored scales [13], and the effect of item bias on validity of PISA [14]. The outcomes of PISA validity turned out to be diverse [9, 15] .

In the aforementioned studies on validating PISA items, testing construct validity is mostly used because PISA does not disclose all the questions in the assessments, it is difficult to testify its content validity and criterion validity such as concurrent, convergent, and discriminant validity through empirical studies [16]. Testifying construct validity of PISA is feasible because the results of different items allow the evaluation on whether students' performance could be incorporated into related theoretical framework or construct. Construct can be viewed as the

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0076

theoretical underpinning that permit researchers to state specific hypotheses about the relationship between abilities and observed behavior [6]. Different constructs could be found in the previous construct validity studies. In particular, Stadler et al. [9] addressed construct validity of collaborative problem-solving (CPS) skills with the construct of CPS performance, students' self-reported collaboration, teacher-reported collaboration, and students' reasoning performance. Stankov et al. [13] used 12 noncognitive scales to examine construct validity of the anchoring method which combined individuals' responses to vignettes and self-rated scores and scrutinized the construct validity of the opportunity to learn (OTL) measure in PISA with the construct of mathematics achievement, opportunity to learn, academic self-concept in mathematics, and ESCS [11].

The construct of this study lies in the relation between metacognition, cognition and self-efficacy. Metacognition is defined by Flavell in terms of its relation to cognition [17], referring to the regulation of one's cognitive activities in learning processes. Self-efficacy beliefs refer to the judgment of confidence in performing academic tasks or succeeding in academic activities. With respect to the relation between metacognition and self-efficacy, it is indicated by the social cognitive theory that students' self-efficacy can influence metacognition, because self-efficacy has an impact on the way people think, feel, act, and regulate their own behavior through cognitive, motivational, affective, and selective processes [18]. Apart from the theories that explicitly draw the relation between metacognition, cognition and self-efficacy, empirical studies also proved strong correlation between metacognition and cognition [19, 20, 21], between metacognition and self-efficacy [22], and among metacognition, cognition and self-efficacy [23, 24], which fortify the construct among metacognition, cognition and self-efficacy.

Among the literature reviewed above, very few focused on the validity of metacognition [5, 7]. The metacognition validated in the previous PISA research was limited to the assessment on metacognition of understanding and remembering and the metacognition of summarizing because it is till PISA 2018 that the metacognition of assessing credibility has been measured. Up until now, no single study has explored the validity of the newly-developed item of metacognition of assessing credibility. Furthermore, as a large scale cross-national assessment, PISA allows the comparison across different participating countries and regions [25], while the discussion on the cross-country comparability of the tested validity is limited [5]. Based on the above-mentioned research gap, this study aimed to measure the validity of metacognition of assessing online information credibility in PISA 2018 and to evaluate the comparability of the validity across 37 OECD countries.

3. Methods

3.1. Data Source and Variables

Released in December 2019, the latest data from the PISA 2018 (URL: http://www.oecd.org/pisa/data/2018database/) were used. In the PISA 2018, the item related to students' metacognitive skill of assessing credibility (item METASPAM) was first measured in the Student Questionnaire on how useful students thought the reading strategy of assessing the quality and the credibility of a piece of text was in reading tasks [1]. Students' level of cognition is indirectly tested in students' reading literacy assessment (item PV1READ), which is designed to measure the possible cognitive approaches of readers to a text [1]. The variable reflecting students' self-efficacy was derived from three items that measured how competent students perceived themselves to be in reading (item SCREADCOMP). These three variables were measured using different methods: the scoring methods for metacognition was based on pair-wise comparisons; students' cognition was inferred from students' reading literacy assessment which were measured using IRT scaling methodology; and students' self-efficacy was derived from three four-point Likert items. As suggested by Bachman [6], validation of

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0076

language tests based on both the abilities to be measured and the facets of the test methods, therefore the distinct test methods of these three variables can diminish the influence of test methods. Since students' individual background is proven to influence their reading performance which need to be controlled, therefore, two control variables were selected, namely student gender (coded as ST004D01T) and the index of economic, social, and cultural status (ESCS). For the purpose of this study, data of 294,527 samples from 37 Organization for Economic co-operation and Development (OECD) countries that have participated in PISA 2018 has been collected, and the demographic information of the samples is presented in Table 1.

Table 1. Demographic information of the samples

Country	N (students)	Female %	N (schools)	SCREADCOMP		METASPAM		ICC
				Mean	SD	Mean	SD	
Australia	14273	7198	763	0.11	0.99	0.19	0.98	0.1808
Austria	6802	3481	291	0.12	1.00	0.32	1.01	0.4757
Belgium	8475	4204	288	-0.01	1.00	-0.25	0.87	0.4679
Canada	22653	11345	821	-0.08	0.99	0.24	1.01	0.1525
Chile	7621	3807	254	-0.32	0.94	-0.18	0.9	0.3821
Colombia	7522	3665	247	-0.28	0.91	-0.17	0.78	0.3938
Czech Republic	7019	3501	333	0.03	0.99	-0.14	0.93	0.5245
Denmark	7657	3841	348	0.08	1.01	0.29	0.9	0.1728
Estonia	5316	2665	230	0.03	0.95	-0.12	0.89	0.2025
Finland	5649	2877	214	0.13	1.02	0.07	0.99	0.0822
France	6308	3230	252	-0.02	1.00	-0.23	0.99	0.5112
Germany	5451	2926	223	0.14	1.00	0.11	0.95	0.5007
Greece	6403	3225	242	0.08	0.95	0.07	0.86	0.3717
Hungary	5132	2527	238	-0.27	0.87	-0.04	0.94	0.5811
Iceland	3296	1640	142	-0.25	0.91	-0.06	1.11	0.0691
Ireland	5577	2800	157	0.19	0.96	0.11	1.00	0.141
Israel	6623	3079	174	-0.26	0.96	0.38	1.02	0.4854
Italy	11785	6105	542	-0.07	0.92	-0.37	0.95	0.4439
Japan	6109	2989	183	0.26	1.03	-0.64	1.04	0.393
Korea	6650	3459	188	-0.3	0.99	-0.19	0.95	0.2611
Latvia	5303	2618	308	-0.02	0.93	-0.29	0.84	0.2121
Lithuania	6885	3508	362	-0.14	0.96	0.33	1.02	0.3827
Luxembourg	5230	2636	44	-0.14	1.01	0.07	1.04	0.2974
Mexico	7299	3473	286	-0.43	0.87	-0.13	0.75	0.3798
Netherlands	4765	2435	156	0.08	0.95	-0.18	0.88	0.5436
New Zealand	6173	3019	192	0.10	0.99	0.05	1.00	0.1654
Norway	5813	2933	251	-0.09	1.02	0.18	1.02	0.0954
Poland	5625	2768	240	-0.05	0.97	-0.14	0.88	0.1835
Portugal	5932	2988	276	-0.02	1.00	-0.25	0.83	0.2877
Slovak Republic	5965	2963	376	-0.23	0.91	-0.44	0.82	0.4496
Slovenia	6401	3408	345	-0.14	0.94	0.03	0.91	0.4844
Spain	35943	17987	1089	-0.01	0.94	-0.12	0.93	0.1544
Sweden	5504	2741	223	0.02	1.02	0.30	1.02	0.1652
Switzerland	5822	3033	228	0.01	0.99	-0.02	0.99	0.3587
Turkey	6890	3494	186	-0.26	0.96	0.02	0.97	0.5717
United Kingdom	13818	6822	471	0.17	0.98	0.14	1.01	0.1778
United States	4838	2462	164	-0.05	0.99	0.25	0.99	0.1708

3.1. Multilevel mediation modeling

As suggested by Bachman [6], the most extensively used approach in construct validation studies is to quantify the correlations among test scores by using quantitative methods, such as factor analysis, causal modeling, and multiple linear regression analysis, etc. Mediation analysis

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0076

is an approach analyzing the underlying relationship between antecedent variable and subsequent variable through one or more mediator. By establishing construct of antecedent variable, subsequent variable and mediator, mediation analysis is often adopted in the studies testing construct validity of PISA [7]. To discover the underlying mechanism through which the independent variable imposed an effect on the dependent variable through intermediary variables, mediation analysis was selected in this study. The PISA data are characterized by the hierarchical feature that student-level data are nested at the school level, which are also nested at the country/region level. Therefore, the intraclass correlation coefficient (ICC) was calculated and confirmed that there was a need for multilevel modeling [26]. Considering the multilevel nature of the data, multilevel mediation analysis model was selected in this study.

Figure 1 shows the model of the multilevel mediation analysis. In this diagram, *X* represents the independent variables of self-efficacy (SCREADCOMP). *Y* represents students' digital reading scores (PV1READ). The relationship between *X* and *Y* is mediated by the mediating variable of the metacognitive process of "assessing credibility" (METASPAM), represented as *M*. C_1 and C_2 represent gender and ESCS, respectively. The categorical variable of gender was converted to a dummy variable, with boys coded as 0 and girls coded as 1. The arrows indicate the direction of the effects.

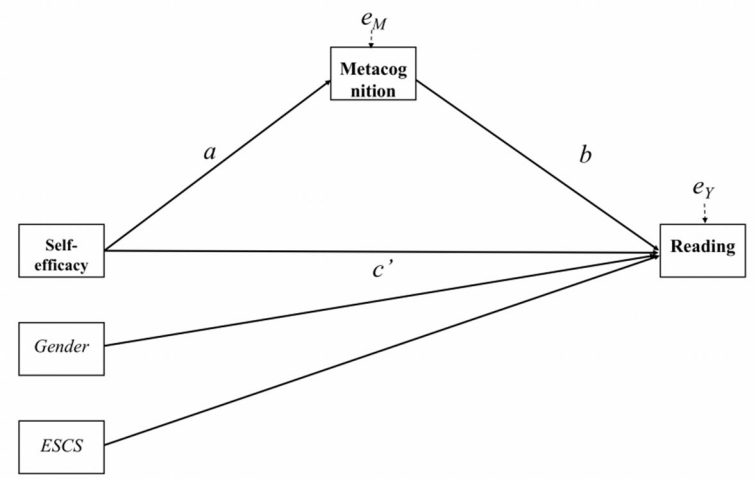

Figure 1. Conceptual diagram of the multilevel mediation model

In mediation analysis, the total direct effect reaches the target without any mediator; in contrast, the indirect effect is expected to pass through the mediating variable on a path from the independent variable to the dependent variable, which can be quantified by multiplying the direct effect of *X* on *M* (*a*) and the direct effect of *M* on *Y* (*b*). The symbol *c'* represents the direct effect of *X* on *Y* holding other variables constant [27]. The total effect of *X* on *Y* (*c*) equals to the value summing the direct (*c'*) and the indirect effects (*a*b*) of *X* on *Y*. The total direct, direct, and indirect effects were all standardized prior to analysis, and the results can be interpreted as standardized mean differences analogous to *Cohen's d*. [27]. To better demonstrate these relations, the following equations are presented according to the framework of mediation analysis [27]:

$$M = i_1 + aX + e_M \quad (1)$$

$$Y = i_2 + c'X + bM + e_{Y1} \quad (2)$$

$$Y = i_3 + cX + e_{Y2} \quad (3)$$

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0076

Where

i_1, i_2, and i_3 are the constants of each regression;

e_M, e_{Y1}, and e_{Y2} are errors in the calculation of M and Y;

a quantifies the relative direct effects of X on the mediator;

b quantifies the direct effect of the mediator on Y when the other variables are controlled;

c quantifies the total effect of X on Y; and

c' quantifies the direct effect of X on Y when the effect of the mediator on Y is controlled, and the the sum of the direct effect (c') and the indirect effect (a*b) of X on Y statistically equals to the total effect of X on Y (c).

In the mediation analysis, a common approach of measuring the effect of the mediator is to quantify the proportion of the total effect that is mediated by the mediator or mediators, the equation of which could be presented as:

$$P_M = a * b/c = a * b/(c' + a * b) \quad (4)$$

It is claimed that the closer PM is closer to 1, the more effect that X on Y mediated by the mediator is observed; in contrast, the closer PM is closer to 0, the less effect of the mediator is. This frequently used effect measurement of calculating the proportion of the total effect is mediated by the intervening variable has been cited in substantive mediation research [29].

3.2. Data Processing

Prior to data processing, the data were preprocessed. Concerning the missing data, imputation with nearest neighbor classification, also known as K-nearest neighbors (KNN), was adopted. The basic logic of KNN imputation is to use the values and the weighted average of the nearest patterns (neighbors) to fill in the target pattern, in which K is the number of neighbors [30]. It has been proven that imputation with KNN is a more robust method for missing value estimation than singular value decomposition and the commonly used row average method [31]. In the current study, KNN imputation was realized in the R package of *DMwR* using the function knnImputation, which utilized the median for numeric variables and the most frequent value for factors to fill in the missing values [32]. Considering the controlled variable of gender was categorical, dummy coding was used to avoid the pitfall of multicollinearity [33]. To ensure that the results of the samples were representative of the population, student weights were processed in R [34], and unbiased population-level parameters were obtained.

Multilevel mediation analysis was performed in R [34] using the structural equation modeling (SEM) approach in latent variable analysis (i.e., lavaan package) [35]. SEM can not only quantify the total, direct, and indirect effects [28] as required in mediation analysis but is also suitable for hierarchical data, with reduced bias in the estimation compared to that seen in other multilevel approaches [29]. In the studies testing construct validity of PISA, SEM is frequently used to unveil the construct among different indicators [9].

4. Results and Discussion

4.1. Total and Direct Effects of Self-efficacy on Digital Reading Performance and the Controlled Effects of Gender and ESCS

As shown in Table 2, the 95% confidence intervals of the total and direct effects of students' self-efficacy in learning on their digital reading performance in all the 37 OECD countries do not include zero, which indicate statistically significant relation between students' self-efficacy

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0076

of reading and their digital reading performance. The unstandardized regression coefficients of the total effects of self-efficacy on digital reading performance range from 0.3683 to 0.4470 in the 37 OECD countries, unveiling that the variance that students' self-efficacy explained to the total variance of students' digital reading is 0.3683 to 0.4470, which reflects medium effect size according to Cohen [36]. As in PISA reading test, students' digital reading performance was designed to reflect their cognitive ability, this result testified significant relation between students' self-efficacy and cognition, which is in line with many other previous studies [18, 23, 24]. This finding contributes to the existing literature with the evidence from large scale international assessment in which the results of 37 OECD countries all show comparable significant relations students' self-efficacy and cognition. To mitigate the influence of students' gender and ESCS background, control variables were included in the mediation model, and the results were presented in Table 2. As shown in the table, the effects of both gender and ESCS on students' digital reading performance were significant in all countries. The coefficients of student gender on digital reading performance were significantly negative, as male students scored significantly lower than female students, but the coefficients of ESCS on digital reading performance were positive, as students with higher ESCS generally had better digital reading performance.

Table 2 Total and direct effects of self-efficacy on digital reading performance and the controlled effects of gender and ESCS.

Country	Total effect			Direct effect			Gender			ESCS		
	B	95% CI		B	95% CI		B	95% CI		B	95% CI	
Australia	0.3683	0.3541	0.3837	0.2821	0.2686	0.2964	-0.1462	-0.1707	-0.1226	0.1897	0.1747	0.2033
Austria	0.3488	0.3278	0.3698	0.2636	0.2443	0.2832	-0.1465	-0.1845	-0.1086	0.2101	0.1919	0.2292
Belgium	0.2559	0.2361	0.2744	0.1936	0.1754	0.2116	-0.1680	-0.2023	-0.1338	0.2988	0.2815	0.3176
Canada	0.3828	0.3710	0.3953	0.3014	0.2903	0.3124	-0.1773	-0.2000	-0.1561	0.1411	0.1297	0.1518
Chile	0.2434	0.2240	0.2630	0.1981	0.1798	0.2172	-0.1267	-0.1661	-0.0902	0.3208	0.3029	0.3418
Colombia	0.2015	0.1784	0.2237	0.1676	0.1476	0.1885	-0.1139	-0.1557	-0.0726	0.3180	0.2997	0.3377
Czech Republic	0.2895	0.2692	0.3120	0.2300	0.2100	0.2510	-0.1807	-0.2158	-0.1434	0.3015	0.2821	0.3212
Denmark	0.3142	0.2937	0.3354	0.2447	0.2265	0.2638	-0.2029	-0.2387	-0.1707	0.2147	0.1963	0.2344
Estonia	0.3261	0.2973	0.3536	0.2664	0.2407	0.2913	-0.1923	-0.2401	-0.1461	0.1626	0.1399	0.1862
Finland	0.3633	0.3359	0.3877	0.2784	0.2549	0.3003	-0.3008	-0.3405	-0.2596	0.1865	0.1655	0.2079
France	0.2384	0.2166	0.2593	0.1755	0.1557	0.1954	-0.1728	-0.2117	-0.1371	0.3353	0.3163	0.3551
Germany	0.3067	0.2823	0.3337	0.2087	0.1863	0.2321	-0.0797	-0.1198	-0.0406	0.2526	0.2287	0.2755
Greece	0.2328	0.2087	0.2558	0.1937	0.1721	0.2159	-0.2740	-0.3178	-0.2306	0.2210	0.2021	0.2445
Hungary	0.3235	0.2995	0.3469	0.2915	0.2673	0.3144	-0.1509	-0.1941	-0.1044	0.3124	0.2915	0.3366
Iceland	0.3870	0.3548	0.4200	0.3199	0.2888	0.3499	-0.3073	-0.3636	-0.2500	0.1358	0.0999	0.1669
Ireland	0.4470	0.4230	0.4700	0.3710	0.3480	0.3940	-0.1810	-0.2250	-0.1410	0.1850	0.1630	0.2060
Israel	0.2148	0.1941	0.2377	0.1872	0.1668	0.2086	-0.2650	-0.3045	-0.2217	0.2827	0.2625	0.3029
Italy	0.3536	0.3358	0.3700	0.2973	0.2815	0.3139	-0.1351	-0.1666	-0.1039	0.1924	0.1760	0.2077
Japan	0.3044	0.2812	0.3293	0.2007	0.1797	0.2224	-0.0496	-0.0889	-0.0093	0.1780	0.1576	0.1993
Korea	0.2780	0.2551	0.3031	0.2230	0.2015	0.2446	-0.1125	-0.1529	-0.0730	0.1568	0.1365	0.1802
Latvia	0.3225	0.2975	0.3509	0.2692	0.2449	0.2951	-0.2609	-0.3087	-0.2164	0.1510	0.1269	0.1746
Lithuania	0.2744	0.2534	0.2960	0.2432	0.2222	0.2636	-0.2627	-0.3021	-0.2247	0.2572	0.2369	0.2788
Luxembourg	0.2909	0.2652	0.3147	0.2183	0.1951	0.2406	-0.1792	-0.2224	-0.1389	0.2605	0.2377	0.2831
Mexico	0.2717	0.2491	0.2925	0.2405	0.2191	0.2610	-0.0875	-0.1281	-0.0469	0.2674	0.2456	0.2864
Netherlands	0.2895	0.2640	0.3172	0.1984	0.1760	0.2224	-0.1748	-0.2224	-0.1318	0.2109	0.1894	0.2351
New Zealand	0.3945	0.3714	0.4170	0.3056	0.2854	0.3267	-0.1818	-0.2181	-0.1410	0.1907	0.1695	0.2101
Norway	0.3499	0.3277	0.3764	0.2651	0.2435	0.2884	-0.3057	-0.3447	-0.2619	0.1424	0.1215	0.1634
Poland	0.3283	0.3035	0.3554	0.2827	0.2594	0.3079	-0.1717	-0.2162	-0.1318	0.2151	0.1883	0.2361
Portugal	0.2614	0.2403	0.2873	0.1935	0.1723	0.2171	-0.1478	-0.1875	-0.1039	0.2208	0.1981	0.2411
Slovak Republic	0.2776	0.2528	0.3024	0.2413	0.2173	0.2652	-0.2294	-0.2712	-0.1875	0.3060	0.2830	0.3271
Slovenia	0.2700	0.2468	0.2928	0.2303	0.2096	0.2522	-0.3443	-0.3827	-0.3017	0.2321	0.2129	0.2549
Spain	0.2507	0.2406	0.2600	0.3128	0.3022	0.3224	-0.1739	-0.1920	-0.1580	0.2056	0.1957	0.2143
Sweden	0.2098	0.1882	0.2310	0.2918	0.2670	0.3147	-0.1975	-0.2371	-0.1533	0.1935	0.1722	0.2143
Switzerland	0.1998	0.1782	0.2230	0.2604	0.2352	0.2848	-0.1633	-0.2005	-0.1169	0.2413	0.2210	0.2642
Turkey	0.1229	0.0982	0.1461	0.1011	0.0776	0.1202	-0.1981	-0.2392	-0.1536	0.2826	0.2615	0.3013
United Kingdom	0.3804	0.3635	0.3944	0.3014	0.2866	0.3153	-0.1754	-0.2003	-0.1481	0.1468	0.1324	0.1608
United States	0.2826	0.2557	0.3115	0.2224	0.1990	0.2478	-0.1744	-0.2235	-0.1319	0.1897	0.1644	0.2137

Note. B is the unstandardized model coefficient, and the 95% CI represents the 95% bias-corrected confidence interval based on the bootstrapping method with 1000 samples.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0076

4.2. Indirect Effect of Self-efficacy on Digital Reading Performance and the Proportion Mediated By Mediator

Table 3 Indirect effect of self-efficacy on digital reading performance and the proportion mediated by mediator.

Country	a path			b path			a*b path			Proportion mediated
	B	95% CI		B	95% CI		B	95% CI		
Australia	0.2157	0.2001	0.2316	0.3997	0.3862	0.4132	0.0862	0.0790	0.0932	0.2341
Austria	0.2012	0.1781	0.2246	0.4232	0.4051	0.4417	0.0852	0.0750	0.0955	0.2442
Belgium	0.1636	0.1412	0.1829	0.3810	0.3632	0.3989	0.0623	0.0532	0.0704	0.2436
Canada	0.2114	0.1982	0.2242	0.3850	0.3745	0.3962	0.0814	0.0757	0.0870	0.2126
Chile	0.1333	0.1106	0.1572	0.3400	0.3204	0.3590	0.0453	0.0373	0.0535	0.1862
Colombia	0.0952	0.0712	0.1169	0.3557	0.3383	0.3744	0.0339	0.0255	0.0419	0.1681
Czech Republic	0.1678	0.1453	0.1919	0.3546	0.3350	0.3748	0.0595	0.0509	0.0683	0.2055
Denmark	0.1630	0.1413	0.1841	0.4260	0.4077	0.4442	0.0694	0.0594	0.0789	0.2210
Estonia	0.1568	0.1273	0.1832	0.3809	0.3598	0.4048	0.0597	0.0489	0.0709	0.1832
Finland	0.2009	0.1764	0.2258	0.4230	0.4025	0.4434	0.0850	0.0743	0.0962	0.2339
France	0.1757	0.1514	0.2011	0.3578	0.3356	0.3771	0.0629	0.0536	0.0728	0.2637
Germany	0.2182	0.1904	0.2434	0.4491	0.4279	0.4699	0.0980	0.0856	0.1102	0.3195
Greece	0.1131	0.0847	0.1353	0.3459	0.3245	0.3663	0.0391	0.0301	0.0477	0.1680
Hungary	0.1214	0.0914	0.1481	0.2633	0.2409	0.2851	0.0320	0.0234	0.0397	0.0988
Iceland	0.1982	0.1635	0.2314	0.3383	0.3107	0.3681	0.0671	0.0551	0.0813	0.1733
Ireland	0.2210	0.1950	0.2450	0.3430	0.3220	0.3630	0.0758	0.0660	0.0866	0.1700
Israel	0.0763	0.0547	0.1023	0.3626	0.3424	0.3822	0.0277	0.0194	0.0367	0.1287
Italy	0.1816	0.1635	0.2002	0.3100	0.2953	0.3265	0.0563	0.0498	0.0627	0.1593
Japan	0.2189	0.1928	0.2437	0.4737	0.4526	0.4946	0.1037	0.0919	0.1164	0.3406
Korea	0.1322	0.1072	0.1555	0.4155	0.3942	0.4330	0.0549	0.0442	0.0651	0.1977
Latvia	0.1535	0.1251	0.1809	0.3475	0.3246	0.3717	0.0533	0.0433	0.0639	0.1654
Lithuania	0.0912	0.0667	0.1151	0.3418	0.3227	0.3637	0.0312	0.0227	0.0394	0.1136
Luxembourg	0.1801	0.1541	0.2073	0.4031	0.3798	0.4235	0.0726	0.0608	0.0842	0.2496
Mexico	0.1057	0.0826	0.1281	0.2949	0.2756	0.3163	0.0312	0.0244	0.0387	0.1147
Netherlands	0.1911	0.1646	0.2208	0.4769	0.4534	0.5010	0.0911	0.0780	0.1053	0.3147
New Zealand	0.2331	0.2092	0.2583	0.3815	0.3617	0.4016	0.0889	0.0789	0.0998	0.2254
Norway	0.2131	0.1866	0.2366	0.3981	0.3783	0.4204	0.0848	0.0736	0.0952	0.2424
Poland	0.1475	0.1204	0.1743	0.3096	0.2890	0.3309	0.0457	0.0368	0.0552	0.1391
Portugal	0.1651	0.0131	0.1919	0.4116	0.3907	0.4356	0.0679	0.0570	0.0799	0.2599
Slovak Republic	0.1264	0.0991	0.1531	0.2868	0.2660	0.3066	0.0363	0.0287	0.0444	0.1306
Slovenia	0.1186	0.0931	0.1432	0.3341	0.3100	0.3535	0.0396	0.0306	0.0481	0.1468
Spain	0.1790	0.1681	0.1896	0.3471	0.3380	0.3554	0.0621	0.0580	0.0660	0.1986
Sweden	0.1842	0.1579	0.2080	0.4454	0.4253	0.4680	0.0820	0.0706	0.0946	0.2811
Switzerland	0.1507	0.1278	0.1769	0.4019	0.3817	0.4235	0.0606	0.0514	0.0717	0.2327
Turkey	0.0605	0.0370	0.0858	0.3603	0.3398	0.3798	0.0218	0.0130	0.0309	0.1774
United Kingdom	0.2036	0.1862	0.2209	0.3881	0.3737	0.4007	0.0790	0.0717	0.0862	0.2078
United States	0.1473	0.1194	0.1799	0.4091	0.3864	0.4332	0.0603	0.0484	0.0735	0.2132

Note. B is the unstandardized model coefficient, and the 95% CI represents the 95% bias-corrected confidence interval based on the bootstrapping method with 1000 samples.

In the Table 3, results of a path reveal the relation between students' self-efficacy in reading and their metacognition of assessing credibility, b path reflects the relation between students' metacognition of assessing credibility and their underlying cognition reflected by the digital reading results. Results of *a*b* path represent the indirect effect that students' self-efficacy impose on students' digital reading performance through the mediation of metacognition of assessing credibility. As indicated by the 95% CI, the unstandardized regression coefficients of a path, *b* path, and *a*b* path are all significant. The coefficients of *a* path range from 0.0605 to 0.2157, which unveil that there is a small to medium effect size of students' self-efficacy in reading imposing on their metacognition of assessing credibility according to Cohen [36]. The statistical significant correlation between students' self-efficacy and metacognition that this study found is consistent with many previous research [18, 22], and is characterized by extending the scope to the particular field of digital reading with the specification of metacognition on assessing online information credibility. The coefficients of b path range from

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0076

0.2633 to 0.4769, which uncover a medium to large effect size of students' metacognition of assessing credibility on their digital reading performance [36]. The significant correlation between metacognition of assessing online information credibility and their cognitive ability revealed by students' digital reading performance provides is congruent with the theoretical underpinning [17] as well as the empirical studies between metacognition and cogniton [19, 20, 21]. To quantify the effect of a mediator, proportion that the total effect is mediated by the mediator is reported in the Table 4, in which the proportion mediated reported in the results ranges from 0.0988 to 0.3406, indicating that 0.0988 to 0.3406 of the the total effect of students' self-efficacy on their digital reading performance is mediated by the metacognition of assessing online credibility across 37 countries. The results further proved that significant correlation could be found among students' self-efficacy, metacognition and cognition, which concurs with previous studies [23, 24]. The findings above jointly prove that the construct of students' self-efficacy in reading, metacognition of assessing online information credibility and their cognitive ability of reading is statistically valid, which further testify the validity of the newly-developed item of measuring students' metacognition of assessing online information credibility and its cross-national comparability.

5. Conclusion and limitations

The objective of the current study is to testify the validity of metacognition of assessing online information credibility in the PISA 2018 and to evaluate the comparability of its validity across 37 OECD countries. In the mediation analysis, the relation between students' self-efficacy in reading and their cognitive ability in reading (represented as *c* path), the relation between students' self-efficacy in reading and their metacogniton of assessing online information credibility (represented as *a* path), the relation between students' metacogniton and their cognitive ability in reading (represented as *b* path), and the relation between students' self-efficacy in reading and their cognitive ability in reading mediated by the metacogniton (represented as *a*b* path) are all statistically significant. The above significant results of all the paths in the mediation analysis jointly manifest that the construct of students' self-efficacy, metacognition and cognition is statistically valid, which is in line with many previous theoretical framework [17, 18] and empirical studies [23, 24]. Since the statistically significant results of students' performance are consistent with the hypothesized constructs of self-efficacy, metacognition and cognition, the construct validity of this study is therefore testified. Indicated by the results across 37 OECD countries, even though different effect sizes were observed across countries, the statistically robust construct could be found in all the selected countries, which corroborates cross-national comparability of the validity of the newly-developed item on measuring the metacognition of assessing online information credibility. The finding that the item of metacognition of assessing unline information credibility in PISA is valid across different countries could facilitate future studies on this particular metacognition and some peripheral studies on how it is related to other learning parameters so that more insightful understanding on this important metacogniton could be achieved. The cross-national comparability of the validity of metacognition of assessing information credibility could also facilitate educators and policy makers worldwide to conduct effective pedagogical measures on students' effective learning.

The current study is by no means without limitations. Since the data of this study are extracted from PISA, the choices of variables to be incorporated in the construct are limited to the existing items which narrows the complexity of the construct that this study adopted. If more items related to the metacognition of assessing online information credibility are available in the following rounds of PISA, it is highly suggested that future studies should testify

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0076

the validity of this particular metacognition with more sophisticated construct which might unveil more valuable conclusions.

References

[1] OECD: PISA 2018 Assessment and Analytical Framework (OECD Publishing, Fance 2019).

[2] J. Coiro and E. Dobler: Exploring the Online Reading Comprehension Strategies Used by Sixth-Grade Skilled Readers to Search for and Locate Information on the Internet, Reading Research Quarterly, Vol. 42 (2007) No. 2, p. 214-257.

[3] J. Abendroth and T. Richter: How to Understand What You don't Believe: Metacognitive Training Prevents Belief-Biases in Multiple Text Comprehension, Learning and Instruction, Vol. 71 (2021), Article 101394.

[4] F. Lang, Y. Kammerer, K. Stürmer and P. Gerjets: Investigating Professed and Enacted Epistemic Beliefs about the Uncertainty of Scientific Knowledge When Students Evaluate Scientific Controversies, European Journal of Psychology of Education, Vol. 36 (2021) No. 1, p. 125-146.

[5] J. Zhou, J. He and D. Lafontaine: Cross-Cultural Comparability and Validity of Metacognitive Knowledge in Reading in PISA 2009: A Comparison of Two Scoring Methods, Assessment in Education: Principles, Policy & Practice, Vol. 27 (2020) No. 6, p. 635-654.

[6] L. F. Bachman: Fundamental Considerations in Language Testing (Oxford University Press, UK 1990).

[7] C. Artelt and W. Schneider: Cross-country Generalizability of the Role of Metacognitive Knowledge in Students' Strategy Use and Reading Competence, Teachers College Record, Vol. 117 (2015) No. 1, Article 010304.

[8] S. Grek: Governing by Numbers: The PISA 'Effect' in Europe, Journal of Education Policy, Vol. 24 (2009) No. 1, p. 23-37.

[9] M. Stadler, K. Herborn, M. Mustafic and S. Greiff: The Assessment of Collaborative Problem Solving in PISA 2015: An Investigation of the Validity of the PISA 2015 CPS Tasks, Computers & Education, Vol. 157 (2020), Article 103964.

[10] M. D. Meier and H. Diefenbach: The OECD Between Political and Scientific Agendas—A Critique of the 2015 PISA Gender Report, Gender and Education, Vol. 32 (2020) No. 5, p. 626-645.

[11] K. Y. Hansen and R. Striethol: Does Schooling Actually Perpetuate Educational Inequality in Mathematics Performance? A Validity Question on the Measures of Opportunity to Learn in PISA, Zdm-Mathematics Education, Vol. 50 (2018) No. 4, p. 643-658.

[12] I. Arffman: Threats to Validity When Using Open-Ended Items in International Achievement Studies: Coding Responses to the PISA 2012 Problem-solving test in Finland, Scandinavian Journal of Educational Research, Vol. 60 (2016) No. 6, p. 609-625.

[13] L. Stankov, J. Lee and M. von Davier: A Note on Construct Validity of The Anchoring Method in PISA 2012, Journal of Psychoeducational Assessment, Vol. 36 (2018) No. 7, p. 709-724.

[14] C. Dzul-Garcia and B. Atar: Investigation of Possible Item Bias on PISA 2015 Science Items Across Chile, Costa Rica and Mexico, Culture Education, Vol. 32 (2020) No.3, p. 470-505.

[15] A. Hanberger: What PISA Intends to and can Possibly Achieve: A critical Programme Theory Analysis, European Educational Research Journal, Vol. 13 (2014) No. 2, p. 167-180.

[16] K. C. Lau: A Critical Examination of PISA's Assessment on Science Literacy, International Journal of Science and Mathematics Education, Vol. 7 (2009) No. 6, p. 1061-1088.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202112_4(12).0076

[17] J. H. Flavell: Metacognition and Cognitive Monitoring: A New Area of Cognitive—Developmental Inquiry, American Psychologist, Vol. 34 (1979) No. 10, p. 906-911.

[18] A. Bandura: Self-efficacy: The Exercise of Control (Freeman and Company, New York 1997).

[19] D. R. Cross and S. G. Paris: Developmental and Instructional Analyses of Children's Metacognition and Reading Comprehension, Journal of Educational Psychology, Vol. 80 (1988) No. 2, p. 131-142.

[20] S. G. Paris and P. Winograd: Promoting Metacognition and Motivation of Exceptional Children, Remedial and Special Education, Vol. 11 (1990) No. 6, p.7-15.

[21] G. Schraw and D. Moshman: Metacognitive Theories, Educational Psychology Review, Vol. 7 (1995) No. 4, p. 351-371.

[22] S. Coutinho: Self-Efficacy, Metacognition, and Performance, North American Journal of Psychology, Vol. 10 (2008) No. 1, p. 165-172.

[23] L. Ma, X. Du, K.-T. Hau and J. Liu: The Association Between Teacher-Student Relationship and Academic Achievement in Chinese EFL Context: A Serial Multiple Mediation Model, Educational Psychology, Vol. 38 (2018) No. 5, p. 687-707.

[24] C. M. Roebers, P. Cimeli, M. Rthlisberger and R. Neuenschwander: Executive Functioning, Metacognition, and Self-perceived Competence in Elementary School Children: An Explorative Study on Their Interrelations and Their Role for School Achievement, Metacognition and Learning, Vol. 7 (2012) No. 3, p. 151-173.

[25] K. Herborn, M. Stadler, M. Mustafic and S. Greiff: The Assessment of Collaborative Problem Solving in PISA 2015: Can Computer Agents Replace Humans? Computers in Human Behavior, Vol. 104 (2020), Article 105624.

[26] T. A. B. Snijders and R. J. Bosker: Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling (2nd ed.) (Sage Publications, Thousand Oaks 2012).

[27] A. F. Hayes and K. J. Preacher: Statistical Mediation Analysis with a Multicategorical Independent Variable, British Journal of Mathematical & Statistical Psychology, Vol. 67 (2014) No. 3, p.451-470.

[28] A. F. Hayes: Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millennium, Communication Monographs, Vol. 76 (2009) No. 4, p.408-420.

[29] L. Juhaňák, J. Zounek, K. Záleská, O. Bárta and K. Vlčková: The Relationship between the Age at first Computer Use and Students' Perceived Competence and Autonomy in ICT Usage: A Mediation Analysis, Computers & Education, Vol. 141 (2019), Article 103614.

[30] O. Kramer: K-Nearest Neighbors. In Dimensionality Reduction with Unsupervised Nearest Neighbors (Springer, Berlin Heidelberg 2013).

[31] O. Troyanskaya, M. Cantor, G. Sherlock, et al: Missing Value Estimation Methods for DNA Microarrays, Bioinformatics, Vol. 17 (2001) No. 6, p. 520-525.

[32] L. Torgo: Data Mining with R: Learning With Case Studies (2nd ed.) (Chapman and Hall/CRC Press, New York 2017).

[33] M. A. Hardy: Regression With Dummy Variables (Sage: Newbury Park 1993).

[34] R Core Team: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Austria 2019).

[35] Y. Rosseel: lavaan: An R Package for Structural Equation Modeling, Journal of Statal Software, Vol. 48 (2012) No. 2, p. 1-36.

[36] J. Cohen: Statistical Power Analysis for the Behavioral Sciences (2nd ed.) (Erlbaum, New Jersey 1988).