

Meningkatkan Validitas Instrumen Literasi Sains PISA: Sebuah Perbandingan Model Sebelum dan Sesudah Revisi dengan CFA

Siti Alifah^{1*)}, Riyan Arthur², Yetti Supriyati³, Munali⁴, & Ibnu Salman⁵

^{1,2,3.} Universitas Negeri Jakarta. ^{4.} Universitas Indraprasta PGRI. ^{5.} Badan Riset dan Inovasi Nasional.

INFO ARTICLES

Key Words:

literasi sains; PISA; CFA; validitas konstruk; pengembangan instrumen.



This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Abstract: This study aims to compare the factor structure of the PISA scientific literacy instrument before and after revision, utilizing Confirmatory Factor Analysis (CFA). The research involved 1,108 respondents consisting of secondary school students from various regions in Indonesia. The PISA scientific literacy instrument employed in this study initially contained 60 items measured across four dimensions of scientific literacy: conceptual understanding, problem-solving, concept application, and scientific reasoning. Through the revision process, several items that did not adequately correspond to the measured constructs were removed to improve model-data fit. The CFA results indicate that the revised model demonstrates a significant improvement in fit compared to the initial model, as evidenced by lower RMSEA and SRMR values and higher Standardized Loading Factors (SLF) for all retained items. These findings suggest that the revised model possesses stronger construct validity and is more adaptive to the local context, without compromising the global relevance of the PISA instrument. This study makes a significant contribution to the development of more accurate and relevant scientific literacy assessment instruments, while providing important insights into how instrument revisions can enhance the quality of scientific literacy measurement at both local and global levels.

Abstrak: Penelitian ini bertujuan untuk membandingkan struktur faktor instrumen literasi sains PISA sebelum dan sesudah dilakukan revisi, dengan menggunakan *Confirmatory Factor Analysis* (CFA). Penelitian ini melibatkan 1.108 responden yang terdiri dari siswa sekolah menengah dari berbagai wilayah di Indonesia. Instrumen literasi sains PISA yang digunakan dalam penelitian ini awalnya mengandung 60 butir soal yang diukur berdasarkan empat dimensi literasi sains: pemahaman konsep, pemecahan masalah, aplikasi konsep, dan penalaran sains. Melalui proses revisi, beberapa butir soal yang tidak sesuai dengan konstruk yang diukur dihapus untuk meningkatkan kesesuaian model dengan data yang diperoleh. Hasil analisis CFA menunjukkan bahwa model revisi menghasilkan peningkatan kecocokan yang signifikan dibandingkan dengan model awal, dengan nilai RMSEA dan SRMR yang lebih rendah serta *Standardized Loading Factor* (SLF) yang lebih tinggi untuk semua butir soal yang tersisa. Temuan ini mengindikasikan bahwa model revisi memiliki validitas konstruk yang lebih kuat dan lebih adaptif terhadap konteks lokal, tanpa mengurangi relevansi instrumen PISA secara global. Penelitian ini memberikan kontribusi signifikan bagi pengembangan instrumen evaluasi literasi sains yang lebih akurat dan relevan, serta memberikan wawasan penting tentang bagaimana revisi instrumen dapat meningkatkan kualitas pengukuran literasi sains pada tingkat lokal maupun global.

Correspondence Address: Gd Bung Hatta Jln. Rawamangun Muka, Jakarta Timur, Indonesia 13220, Indonesia; e-mail: siti.alifah2005@yahoo.co.id.

How to Cite (APA 6th Style): Alifah, S., Arthur, R., Supriyati, Y., Munali, & Salman, I. (2026). Meningkatkan Validitas Instrumen Literasi Sains PISA: Sebuah Perbandingan Model Sebelum dan Sesudah Revisi dengan CFA. *Prosiding Diskusi Panel Nasional Pendidikan Matematika*, 647-660.

Copyright: Siti Alifah, Riyan Arthur, Yetti Supriyati, Munali, & Ibnu Salman. (2026)

PENDAHULUAN

Literasi sains adalah kemampuan individu untuk memahami konsep-konsep ilmiah, mengevaluasi informasi berbasis bukti dan menggunakan pengetahuan ilmiah dalam kehidupan sehari-hari, serta untuk membuat keputusan yang berbasis pada bukti ilmiah. Kemampuan ini tidak hanya mencakup pengetahuan teoretis, tetapi juga keterampilan praktis seperti berpikir kritis, pemecahan masalah, dan kemampuan untuk mengevaluasi informasi ilmiah secara reflektif. Dalam konteks global, literasi sains menjadi indikator penting bagi kesiapan generasi muda menghadapi tantangan masa depan yang semakin kompleks dan berbasis teknologi (OECD, 2019).

Teknologi memberi siswa kesempatan untuk melakukan eksplorasi mandiri, memungkinkan siswa untuk memperoleh pengetahuan lebih dalam tanpa tergantung sepenuhnya pada informasi yang diberikan oleh guru. Teknologi dianggap sebagai alat yang mendukung eksplorasi ilmiah yang lebih dalam dengan memberikan siswa kesempatan untuk berinteraksi dengan konten sains secara lebih fleksibel dan dinamis (Premthaisong & Srisawasdi, 2024). Oleh karena itu, evaluasi literasi sains di tingkat internasional, seperti yang dilakukan oleh Program for International Student Assessment (PISA), memberikan gambaran menyeluruh tentang kualitas pendidikan sains di berbagai negara dan membantu merumuskan kebijakan pendidikan yang berbasis bukti (Fensham, 2013; Sjøberg & Jenkins, 2020).

Sebagai instrumen penilaian global, PISA berfungsi untuk mengukur kompetensi literasi sains siswa di lebih dari 80 negara, dengan fokus pada empat dimensi utama: konteks, pengetahuan, kompetensi, dan sikap ilmiah. Namun, penelitian menunjukkan bahwa instrumen yang digunakan dalam pengukuran literasi sains, meskipun telah dirancang secara internasional, sering kali menghadapi tantangan dalam hal validitas konstruk dan kesesuaian dengan konteks lokal (OECD, 2019; Sadler & Zeidler, 2009). Validitas dan reliabilitas instrumen sangat penting untuk memastikan bahwa hasil pengukuran secara akurat mencerminkan kemampuan siswa (Manz et al., 2025; Peeters & Harpe, 2020). Oleh karena itu, revisi dan pengembangan instrumen literasi sains berdasarkan evaluasi data empiris menjadi langkah penting untuk meningkatkan kualitas pengukuran yang lebih tepat dan relevan di setiap konteks pendidikan (Zetterqvist & Bach, 2023; Zhang et al., 2023).

Seiring dengan perkembangan kurikulum dan kebutuhan pendidikan abad ke-21, instrumen literasi sains harus mampu mencerminkan keterampilan berpikir kritis, kreativitas, dan kemampuan problem solving siswa secara komprehensif. Evaluasi struktur faktor instrumen menggunakan metode *Confirmatory Factor Analysis* (CFA) menjadi pendekatan penting dan efektif untuk mengevaluasi dan memperbaiki struktur faktor instrumen pengukuran pendidikan. CFA memungkinkan peneliti untuk menguji kecocokan model antara struktur faktor yang ada dan data yang diperoleh, serta untuk mengidentifikasi butir-butir soal yang tidak sesuai dengan konstruk yang dimaksud (Han & Abdul Rahman, 2025; Rattanachaihada et al., 2025). Dengan membandingkan struktur faktor instrumen literasi sains PISA sebelum dan sesudah revisi, peneliti dapat mengevaluasi sejauh mana perubahan tersebut memperbaiki kualitas instrumen, meningkatkan konsistensi internal, serta memastikan bahwa instrumen dapat digunakan secara lebih valid dalam berbagai konteks pendidikan dan lebih adaptif terhadap konteks lokal (Nurhayati et al., 2024). Hasil dari analisis ini tidak hanya memberikan wawasan tentang validitas instrumen, tetapi juga dapat membantu dalam penyesuaian instrumen agar lebih sesuai dengan kebutuhan pendidikan lokal tanpa mengurangi aspek globalnya (Abu Bakar et al., 2023; Jomtarak et al., 2019; Suwono et al., 2022).

Penelitian ini bertujuan untuk membandingkan struktur faktor instrumen literasi sains PISA sebelum dan sesudah revisi menggunakan *Confirmatory Factor Analysis* (CFA). Melalui CFA, penelitian ini mengevaluasi sejauh mana revisi instrumen dapat meningkatkan kesesuaian antara struktur faktor yang ada dan data yang diperoleh. Fokus utama penelitian adalah untuk mengidentifikasi dan memperbaiki butir-butir soal yang tidak sesuai dengan konstruk yang dimaksud, serta untuk meningkatkan validitas dan reliabilitas pengukuran literasi sains secara lebih akurat. Selain itu, penelitian ini bertujuan untuk memastikan bahwa instrumen yang direvisi dapat digunakan secara efektif dalam berbagai konteks pendidikan yang beragam, mengingat perbedaan budaya dan

kurikulum di negara-negara peserta PISA. Hasil dari analisis ini diharapkan dapat memberikan kontribusi terhadap pengembangan instrumen evaluasi pendidikan yang lebih adaptif dan relevan baik di tingkat global maupun lokal.

Penelitian ini memberikan wawasan yang berharga mengenai efektivitas revisi instrumen literasi sains dalam meningkatkan kualitas pengukuran yang lebih akurat dan relevan. Dengan mengevaluasi struktur faktor instrumen sebelum dan sesudah revisi menggunakan *Confirmatory Factor Analysis* (CFA), penelitian ini menunjukkan bagaimana perubahan pada butir-butir soal dapat memperbaiki validitas dan reliabilitas instrumen, sehingga menghasilkan pengukuran yang lebih konsisten dan tepat. Selain itu, penelitian ini memberikan kontribusi signifikan terhadap pengembangan instrumen evaluasi pendidikan di tingkat internasional dengan menawarkan pendekatan yang dapat diterapkan untuk menyempurnakan instrumen yang digunakan dalam berbagai konteks budaya dan kurikulum. Hasil dari penelitian ini diharapkan dapat memperkaya literatur yang ada dan menjadi referensi untuk penelitian lebih lanjut dalam pengembangan instrumen evaluasi pendidikan yang lebih adaptif dan valid secara global.

Penelitian ini memberikan kebaruan dengan memperbaiki instrumen literasi sains PISA menggunakan *Confirmatory Factor Analysis* (CFA) untuk membandingkan struktur faktor sebelum dan sesudah revisi. Penghapusan butir soal dengan muatan faktor rendah meningkatkan validitas dan reliabilitas instrumen, menjadikannya lebih adaptif terhadap konteks lokal tanpa mengurangi relevansi global. Temuan ini berkontribusi dalam pengembangan instrumen evaluasi literasi sains yang lebih akurat, konsisten, dan sesuai dengan kebutuhan pendidikan di berbagai negara.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan analisis data yang menerapkan *Confirmatory Factor Analysis* (CFA) untuk menguji validitas instrumen literasi sains. CFA dipilih karena kemampuannya untuk menguji struktur faktor yang telah ditentukan sebelumnya, serta mengevaluasi sejauh mana data yang diperoleh sesuai dengan model teoretis yang ada. Desain penelitian yang digunakan adalah eksperimen komparatif, yang bertujuan untuk membandingkan struktur faktor instrumen literasi sains sebelum dan sesudah revisi. Dalam desain ini, peneliti mengumpulkan data dari instrumen yang belum direvisi dan yang telah direvisi, kemudian menganalisis perbedaan dalam validitas hasil pengukuran. Dengan pendekatan ini, diharapkan dapat diperoleh pemahaman yang lebih baik mengenai peningkatan kualitas instrumen dan efektivitas perubahan yang dilakukan dalam meningkatkan pengukuran literasi sains.

Penelitian ini melibatkan total 823 siswa SMP dari Kabupaten Indramayu (480 siswa), Kabupaten Sukabumi (150 siswa), dan Kota Bekasi (193 siswa) sebagai sampel penelitian yang diambil pada semester genap tahun ajaran 2024/2025. Sampel tersebut dipilih untuk mewakili variasi geografis dan demografis yang ada di wilayah Jawa Barat, sehingga diharapkan hasil penelitian dapat menggambarkan kondisi yang lebih representatif. Instrumen yang digunakan dalam penelitian ini adalah instrumen literasi sains berbasis PISA yang terdiri dari dua versi: sebelum revisi, instrumen tersebut terdiri dari 60 butir soal yang mengukur berbagai aspek literasi sains, sementara setelah revisi, instrumen ini terdiri dari 55 butir soal yang disesuaikan untuk meningkatkan validitas dan reliabilitas pengukuran. Perubahan jumlah soal ini dilakukan untuk memperbaiki kualitas pengukuran dan memastikan bahwa instrumen lebih sesuai dengan konteks lokal dan tujuan pengukuran literasi sains yang lebih tepat.

Analisis data dalam penelitian ini menggunakan *Confirmatory Factor Analysis* (CFA) untuk menguji struktur faktor instrumen literasi sains yang digunakan. CFA dipilih karena kemampuannya untuk mengevaluasi sejauh mana data yang diperoleh sesuai dengan model teoretis yang telah ditentukan sebelumnya, serta untuk menguji validitas konstruk dari instrumen yang digunakan. Dalam penelitian ini, dilakukan perbandingan *goodness of fit* antara model struktur faktor instrumen literasi sains sebelum dan sesudah revisi. Perbandingan ini bertujuan untuk menilai sejauh mana perubahan yang dilakukan pada instrumen berkontribusi terhadap peningkatan kesesuaian model

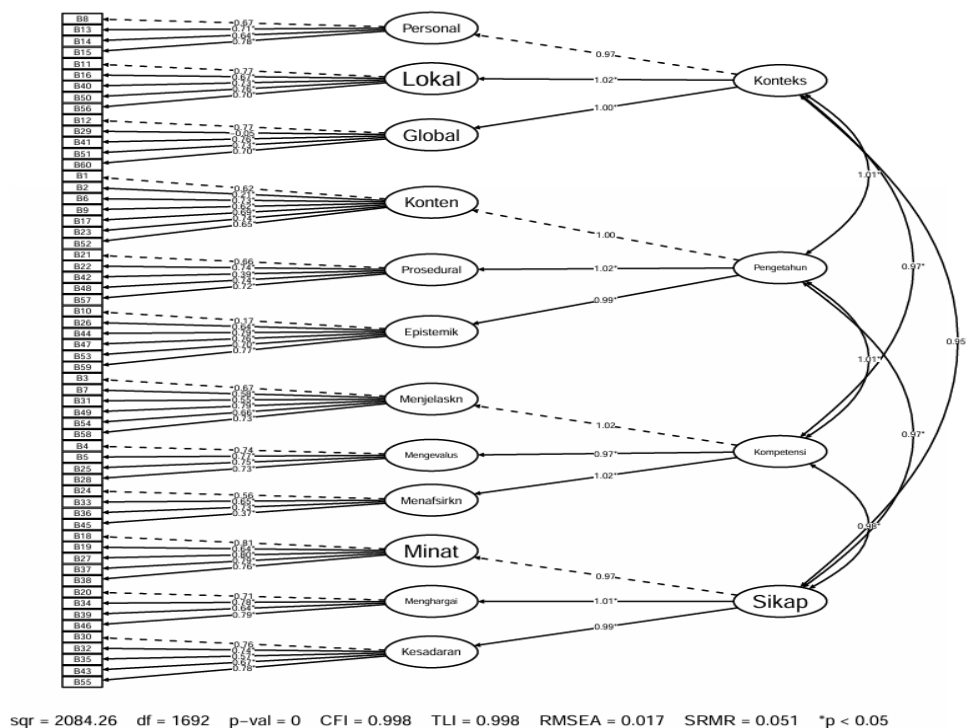
dengan data, serta untuk memastikan bahwa instrumen yang telah direvisi memberikan hasil yang lebih valid dan reliabel. Indikator *goodness of fit* seperti Chi-Square, RMSEA, CFI, dan TLI digunakan untuk mengukur kesesuaian antara model yang diusulkan dan data yang dikumpulkan, sehingga memberikan pemahaman yang lebih baik tentang efektivitas revisi instrumen.

HASIL

Hasil Analisis Faktor Konfirmatori / *Confirmatory Factor Analysis* (CFA)

A. Hasil CFA Model Pengukuran Awal

Model pengukuran awal dikembangkan berdasarkan konstruk konseptual literasi sains siswa SMP, yang mengacu pada empat dimensi utama: konteks, pengetahuan, kompetensi, dan sikap. Model pengukuran awal (konseptual) literasi sains siswa SMP disajikan pada gambar 2 berikut:



Gambar 1. Model Pengukuran Awal Literasi Sains

Sumber: Diolah dari Data Hasil Penelitian, 2025.

Analisis CFA tahap awal menunjukkan bahwa sebagian besar butir memiliki nilai muatan faktor terstandar (*Standardized Loading Factor/SLF*) yang cukup baik, yaitu diatas 0,5. Namun, terdapat lima butir yang memiliki nilai SLF < 0,40, sebagaimana ditunjukkan pada Tabel 2 berikut:

Tabel 1. Butir dengan SLF < 0,40 dalam Model Awal

No	Kode Butir	Dimensi	SLF	Keterangan
1	B2	Pengetahuan Konten	0.214	Lemah
2	B10	Pengetahuan Epistemik	0.172	Lemah
3	B29	Konteks – Global/Internasional	-0.048	Korelasi negatif
4	B42	Pengetahuan Prosedural	0.394	Hampir tidak valid
5	B45	Kompetensi – Menafsirkan data	0.373	Lemah

Sumber: Diolah dari Data Hasil Penelitian, 2025.

Walaupun indeks kecocokan model (fit indices) secara keseluruhan sudah baik (CFI dan TLI = 0,998; RMSEA = 0,017; SRMR = 0,051), keberadaan butir-butir dengan SLF rendah berpotensi menurunkan validitas konstruk serta memperlemah konsistensi internal antar indikator dalam masing-masing dimensi.

Kecocokan Model Pengukuran dengan Data Empiris

Gambar 4.1 menunjukkan beberapa indeks kecocokan model (*Goodness of Fit*) seperti CFI = 0,998 ($> 0,90$), TLI = 0,998 ($> 0,90$), RMSEA = 0,017 ($< 0,08$), dan SRMR = 0,051 ($< 0,08$).

Kecocokan model yang baik diindikasikan oleh beberapa parameter: CFI $\geq 0,90$ (Cho et al., 2020; Sathyanarayana & Mohanasundaram, 2024); TLI $\geq 0,90$ (Shi et al., 2018; Xia & Yang, 2019); RMSEA $< 0,06$ hingga 0,08 (Lai & Green, 2016; Schreiber et al., 2006; Shi et al., 2022); dan SRMR $\leq 0,08$ (Niemand & Mai, 2018; Pavlov et al., 2021). Hasil analisis CFA mengindikasikan bahwa model pengukuran awal telah memenuhi kriteria kecocokan model yang direkomendasikan. Ini menunjukkan bahwa model pengukuran memiliki kecocokan yang baik dengan data empiris, yang berarti juga bahwa model CFA yang digunakan sangat fit terhadap data empiris. Dengan demikian, model pengukuran yang diuji mampu merepresentasikan data secara memadai, dan struktur faktor yang diasumsikan sesuai dengan data yang diamati. Dengan kata lain, struktur faktor dari instrumen literasi sains sesuai dengan desain teoritis yang dirancang.

Muatan Faktor (*Loading Factor*)

Model pengukuran awal memiliki 60 butir pengukuran (variabel manifest/observasi) dengan muatan faktor beragam terhadap variabel latennya. Tabel 3 menyajikan muatan faktor terstandar (*standardized loading factor*) dari semua butir terhadap variabel latennya.

Tabel 2. Muatan faktor terstandar model pengukuran awal

Variabel Laten	Indikator	Butir	SLF	z-value	p-value
1. Dimensi Konteks (Context)	Personal	B13	0.715	14.805	0.000
		B14	0.643	14.432	0.000
		B8	0.670		
		B15	0.778	16.676	0.000
	Lokal / Nasional	B11	0.773		
		B40	0.728	21.846	0.000
		B56	0.703	21.581	0.000
		B16	0.670	17.560	0.000
		B50	0.757	22.185	0.000
	Global / Internasional	B29	-0.048	-0.904	0.366
		B12	0.770		
		B41	0.761	22.329	0.000
		B51	0.727	20.268	0.000
		B60	0.702	17.914	0.000
2. Dimensi Pengetahuan (Knowledge)	Pengetahuan Konten	B1	0.619		
		B17	0.686	14.983	0.000
		B2	0.214	4.638	0.000
		B6	0.732	15.778	0.000
		B9	0.617	12.973	0.000
		B23	0.744	16.210	0.000
	Pengetahuan Prosedural	B52	0.652	14.732	0.000
		B21	0.663		
		B22	0.743	16.980	0.000
		B42	0.394	8.899	0.000
		B48	0.742	17.838	0.000
		B57	0.716	17.207	0.000
	Pengetahuan	B10	0.172		

3. Dimensi Kompetensi (Competencies)	Epistemik	B26	0.645	3.261	0.001
		B44	0.786	3.303	0.001
		B47	0.759	3.308	0.001
		B53	0.698	3.331	0.001
		B59	0.767	3.307	0.001
	Menjelaskan fenomena secara ilmiah	B3	0.668		
		B7	0.578	13.688	0.000
		B31	0.546	13.088	0.000
		B54	0.660	16.490	0.000
		B58	0.734	20.008	0.000
		B49	0.786	20.522	0.000
		B28	0.728	21.713	0.000
		B25	0.749	20.882	0.000
		B4	0.742		
		B5	0.771	21.916	0.000
	Menafsirkan data dan bukti secara ilmiah	B45	0.373	7.574	0.000
		B24	0.558		
		B33	0.648	13.061	0.000
		B36	0.729	13.934	0.000
		B37	0.791	26.471	0.000
4. Dimensi Sikap (Attitudes)	Minat terhadap sains dan teknologi	B27	0.798	28.556	0.000
		B38	0.757	26.687	0.000
		B18	0.808		
		B19	0.642	16.149	0.000
		B20	0.708		
	Menghargai pendekatan ilmiah terhadap penyelidikan	B39	0.641	15.383	0.000
		B34	0.778	19.981	0.000
		B46	0.792	20.110	0.000
		B30	0.755		
	Kesadaran terhadap masalah lingkungan	B32	0.738	21.090	0.000
		B35	0.572	14.519	0.000
		B43	0.669	18.849	0.000
		B55	0.784	25.067	0.000

Sumber: Diolah dari Data Hasil Penelitian, 2025.

Tabel 3 mengindikasikan bahwa muatan faktor terstandar (*standardized loading factor*) adalah merentang dari -0,048 hingga 0,808. Muatan faktor terstandar menunjukkan seberapa kuat suatu butir menjelaskan variabel laten yang diukur. Butir-butir dengan loading faktor tinggi menunjukkan validitas indikator yang kuat. Semakin tinggi muatan faktornya, semakin baik butir tersebut mencerminkan variabel laten yang diukurnya. Dalam konteks pengembangan instrumen, literatur merekomendasikan muatan faktor sekurang-kurangnya 0,40 dapat diterima untuk memastikan validitas konstruk (Helfrich et al., 2007; Lorenzo-Seva, 2025; Watson, 2017).

Pada Dimensi 4 (Sikap), semua butir memiliki muatan faktor $> 0,40$. Pada Dimensi 1 (Konteks) terdapat satu butir yang memiliki muatan faktor $< 0,40$ (yaitu butir 29). Pada Dimensi 2 (Pengetahuan) terdapat tiga butir yang memiliki muatan faktor $< 0,40$ (yaitu butir 2, 10, dan 42); sedangkan pada Dimensi 3 (Kompetensi) terdapat satu butir yang memiliki muatan faktor $< 0,40$ (yaitu butir 45). Butir-butir dengan muatan faktor $< 0,40$ dapat dipertahankan dalam model pengukuran jika terdapat justifikasi teoretis yang kuat dan kecocokan model masih terpenuhi (Evans & Napier-Munn, 2013; Pavlů et al., 2006).

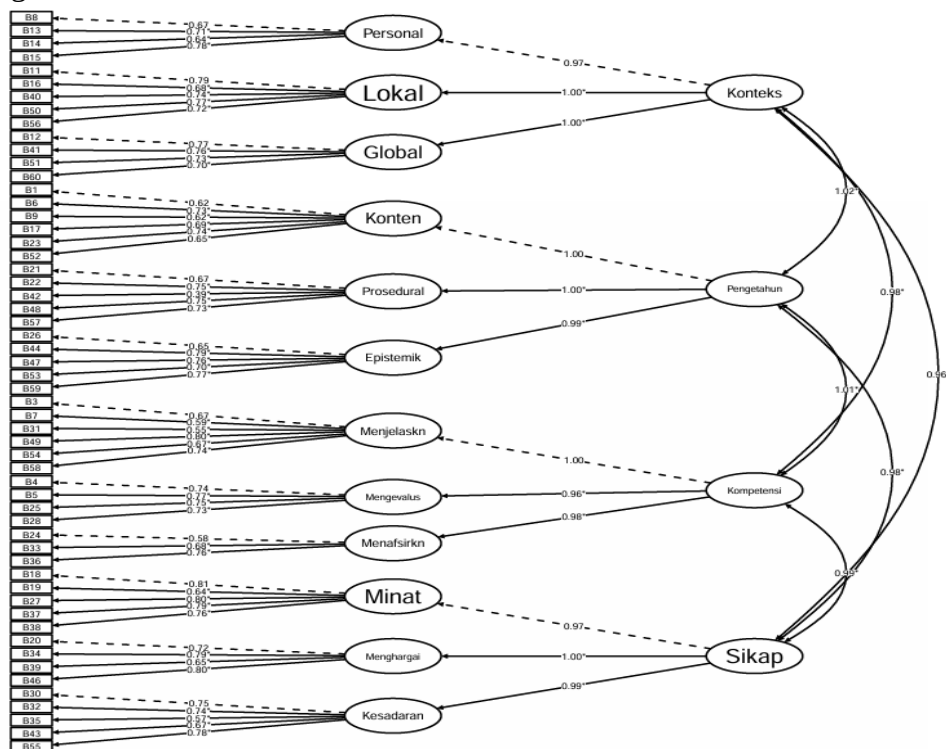
Terkonfirmasi dari 60 butir terdapat lima butir (B2, B10, B29, B42, dan B45) memiliki *loading factor* sangat rendah ($< 0,04$), bahkan ada satu butir yang memiliki *loading factor* negatif, yaitu B29 dengan *loading factor* -0,048 dan muatan faktor tidak signifikan ($p > 0,05$), yaitu 0,366 yang menunjukkan bahwa kontribusinya terhadap konstruk yang diukur hampir tidak ada. Secara statistik, nilai loading seperti ini menunjukkan hubungan yang lemah atau tidak signifikan antara butir dan faktor, sehingga butir tersebut tidak merepresentasikan konstruk secara valid (Chao et al., 2014; Douglas L. Inman, 1952; Inga et al., 2021; Koskenniska et al., 2020). Lebih lanjut Baumgartner et al. (2018) dan Yoon et al. (2020) menjelaskan bahwa butir

soal dengan nilai *loading factor* yang sangat rendah atau negatif (misalnya dibawah 0,40) bisa terjadi karena redaksi soal yang membingungkan (ambigu), penempatan dimensi yang kurang tepat, atau adanya kesalahpahaman dari responden. Menurut Baumgartner et al. (2018), hal ini disebut *misresponse*, yaitu ketika responden menjawab tidak konsisten karena kesulitan memahami isi soal. Kesalahan ini bisa muncul karena kata-kata yang dibalik, kalimat negatif, atau konsep yang berlawanan dengan item lain. Selain karena kurangnya perhatian, masalah ini juga bisa muncul karena responden tidak mampu memahami item dengan benar. Akibatnya, nilai *loading factor* menjadi rendah dan kualitas pengukuran instrumen pun menurun. Hal ini diperkuat Yoon et al. (2020) yang mengonfirmasi bahwa tidak hanya ambiguitas dan miskonsepsi yang berdampak pada kualitas faktor, tetapi juga menunjukkan bahwa item-item dengan penulisan negatif secara sistematis mengganggu struktur faktor instrumen.

Berdasarkan tabel 3, pada pengukuran awal literasi sains tersebut, terkonfirmasi sebagian besar item menunjukkan signifikansi statistik ($p < 0.05$), yang mengindikasikan kontribusi nyata terhadap faktor masing-masing. Hal ini memperkuat evidensi bahwa butir-butir tersebut mengukur aspek literasi sains yang sesuai dengan konstruk teoritisnya. Secara implikasi teoritis dan praktis, instrumen ini mampu mengidentifikasi domain kognitif dan afektif dalam literasi sains secara komprehensif. Korelasi antar faktor yang ditunjukkan melalui model *correlated factors* menunjukkan bahwa aspek literasi sains saling berhubungan secara teoritis dan empiris, sehingga pendekatan integratif dalam pembelajaran sangat tepat digunakan. Temuan ini dapat digunakan untuk pengembangan lebih lanjut pada skala nasional maupun untuk *vertical equating* lintas jenjang pendidikan.

B. Hasil CFA Perbaikan Model Pengukuran

Perbaikan model dilakukan untuk mengetahui apakah mengeluarkan butir-butir dengan muatan faktor $< 0,40$ dapat meningkatkan kecocokan model pengukuran dengan data empiris. Gambar 3 menyajikan model pengukuran setelah mengeluarkan butir-butir dengan muatan faktor kurang dari 0,40.



Chi-sqr = 1791.3 df = 1470 p-val = 0 CFI = 0.998 TLI = 0.998 RMSEA = 0.016 SRMR = 0.049 * $p < 0.05$

Gambar 2. Model pengukuran revisi literasi sains

Sebagai langkah perbaikan, kelima butir tersebut dikeluarkan dari model. Hasil CFA ulang terhadap model yang telah direvisi menunjukkan peningkatan kualitas pengukuran, sebagaimana ditunjukkan dalam Tabel 4.

Tabel 3. Perbandingan Indeks Kecocokan Model CFA Sebelum dan Sesudah Revisi

Indikator Fit Model	Model Awal	Model Revisi	Interpretasi
Chi-square	2084.26	1791.30	Menurun, lebih baik
Derajat kebebasan (df)	1692	1470	Wajar turun karena pengurangan item
CFI	0.998	0.998	Sangat baik
TLI	0.998	0.998	Sangat baik
RMSEA	0.017	0.016	Lebih kecil, lebih baik
SRMR	0.051	0.049	Lebih kecil, mendekati ideal
SLF terendah	0.172	0.549	Meningkat signifikan

Sumber: Diolah dari Data Hasil Penelitian, 2025.

PEMBAHASAN

Penghapusan lima butir dengan muatan faktor $< 0,40$ berkontribusi positif terhadap perbaikan validitas konstruk instrumen. Secara substansi, indikator yang dikeluarkan memiliki beberapa kelemahan:

- B29 menyajikan konteks global yang kemungkinan tidak familiar bagi siswa SMP di lingkungan lokal.
- B2 dan B10 menunjukkan kontribusi yang sangat lemah terhadap konstruk pengetahuan, terutama dalam aspek konten dan epistemik.
- B42 dan B45 menunjukkan bahwa siswa mengalami kesulitan dalam memahami pendekatan prosedural dan interpretasi data dalam format soal yang diberikan.

Model hasil revisi menghasilkan nilai SLF yang semuanya $\geq 0,54$, yang berarti bahwa setiap indikator kini memiliki kontribusi signifikan terhadap variabel laten yang diukur. Selain itu, perbandingan antar dimensi menunjukkan korelasi positif yang kuat namun tetap berada dalam batas normal, yang menunjukkan tidak adanya gejala multikolinearitas.

Perbaikan model CFA ini memberikan beberapa implikasi penting:

1. Secara teoretis, instrumen ini lebih representatif dalam mengukur literasi sains berdasarkan kerangka kerja PISA.
2. Secara empirik, penghapusan butir bermasalah meningkatkan kekonsistenan indikator dan akurasi estimasi parameter laten.
3. Secara praktis, hasil ini menjadi dasar untuk menyempurnakan alat asesmen diagnostik literasi sains di jenjang SMP.

Model pengukuran hasil revisi terbukti lebih baik dibandingkan model awal dari sisi muatan faktor maupun indeks kecocokan model. Instrumen hasil revisi memiliki validitas konstruk yang lebih kuat dan dapat dijadikan dasar untuk pelaksanaan uji lanjutan pada skala populasi yang lebih besar atau dalam penelitian lintas wilayah.

Tabel 4 menunjukkan bahwa CFI dan TLI tetap, sedangkan RMSEA dan SRMR menurun, mengindikasikan bahwa indeks kecocokan model membaik setelah mengeluarkan butir-butir dengan muatan faktor $< 0,40$. Tabel 4 juga menyajikan muatan faktor terstandar dari 55 butir pengukuran tersisa terhadap variabel latennya. Tabel 4 menunjukkan bahwa muatan faktor terstandar merentang dari 0,549 hingga 0,808. Semua muatan faktor juga terkonfirmasi signifikan ($p < 0,05$). Ini mengindikasikan bahwa semua butir berkontribusi signifikan dalam mengukur variabel latennya.

Tabel 4. Muatan faktor terstandar model pengukuran perbaikan

Variabel Laten	Indikator	Butir	SLF	z-value	p-value
1. Dimensi Konteks (Context)	Personal	B13	0.715	14.789	0.000
		B14	0.644	14.437	0.000
		B8	0.669		
		B15	0.778	16.681	0.000
	Lokal / Nasional	B11	0.786		
		B40	0.740	21.841	0.000
		B56	0.715	21.599	0.000
		B16	0.682	17.549	0.000
		B50	0.769	22.140	0.000
	Global / Internasional	B12	0.769		
		B41	0.761	22.325	0.000
		B51	0.728	20.303	0.000
		B60	0.701	17.889	0.000
2. Dimensi Pengetahuan (Knowledge)	Pengetahuan Konten	B1	0.620		
		B17	0.686	14.981	0.000
		B6	0.733	15.795	0.000
		B9	0.617	12.982	0.000
		B23	0.743	16.199	0.000
		B52	0.652	14.738	0.000
	Pengetahuan Prosedural	B21	0.673		
		B22	0.755	17.005	0.000
		B48	0.753	17.839	0.000
		B57	0.727	17.235	0.000
	Pengetahuan Epistemik	B26	0.645		
		B44	0.786	17.002	0.000
		B47	0.758	17.095	0.000
		B53	0.698	15.169	0.000
		B59	0.766	17.018	0.000
3. Dimensi Kompetensi (Competencies)	Menjelaskan fenomena secara ilmiah	B3	0.675		
		B7	0.587	13.734	0.000
		B31	0.549	12.980	0.000
		B54	0.669	16.527	0.000
		B58	0.743	19.987	0.000
	Mengevaluasi dan merancang penyelidikan ilmiah	B49	0.796	20.546	0.000
		B28	0.728	21.734	0.000
		B25	0.749	20.877	0.000
		B4	0.742		
		B5	0.771	21.911	0.000
	Menafsirkan data dan bukti secara ilmiah	B24	0.582		
		B33	0.676	13.099	0.000
		B36	0.760	13.952	0.000
4. Dimensi Sikap (Attitudes)	Minat terhadap sains dan teknologi	B37	0.791	26.465	0.000
		B27	0.797	28.498	0.000
		B38	0.756	26.733	0.000
		B18	0.808		
		B19	0.644	16.211	0.000
	Menghargai pendekatan ilmiah terhadap penyelidikan	B20	0.716		
		B39	0.648	15.453	0.000
		B34	0.785	20.037	0.000
		B46	0.799	20.128	0.000
	Kesadaran terhadap masalah lingkungan	B30	0.755		
		B32	0.739	21.046	0.000
		B35	0.572	14.506	0.000
		B43	0.668	18.796	0.000
		B55	0.784	25.025	0.000

Ket: SLF = *Standardized Loading Factor*

Sumber: Diolah dari Data Hasil Penelitian, 2025.

Model Awal (Sebelum Penghapusan Butir)

Model awal (sebelum penghapusan butir), terdapat 5 butir dengan *loading factor* $< 0,40$, yaitu B2 (0,214), B10 (0,172), B29 (-0,048), B42 (0,394), dan B45 (0,373). Nilai SLF $< 0,40$ menunjukkan bahwa butir tersebut memberikan kontribusi lemah terhadap konstruk yang diukur. Bahkan nilai SLF negatif (seperti B29) menandakan korelasi terbalik yang secara teoritis dan empiris sangat bermasalah, menunjukkan kemungkinan miskonsepsi siswa atau ketidakcocokan konteks lokal terhadap stimulus soal. Butir-butir tersebut dapat menurunkan validitas konstruk / konsistensi internal konstruk karena kontribusi yang lemah terhadap dimensi laten, juga dapat menyebabkan bias dalam estimasi hubungan antar indikator dan mengganggu konvergensi model, terutama jika jumlah sampel terbatas.

Model Revisi (Setelah Penghapusan 5 Butir)

Pada model revisi (setelah penghapusan butir), butir yang memiliki *loading factor* rendah telah dihapus. *Loading factor* setiap indikator meningkat konsistensinya ($> 0,56$), menunjukkan struktur faktor yang lebih stabil dan valid. Model tetap memiliki fit yang sangat baik, dan yang lebih penting: parsimonious (lebih sederhana dan efektif). Nilai korelasi antardimensi menjadi lebih tajam, misalnya hubungan Kompetensi $\leftrightarrow$ Sikap = $0,52^*$ ($*p < 0,05$), menjadi lebih terlihat secara statistik dan substantif.

Model awal (sebelum penghapusan butir), terdapat 5 butir dengan *loading factor* $< 0,40$, yaitu B2 (0,214), B10 (0,172), B29 (-0,048), B42 (0,394), dan B45 (0,373). Butir-butir tersebut dapat menurunkan validitas konstruk karena kontribusi yang lemah terhadap dimensi laten, juga dapat menyebabkan distorsi estimasi parameter dan mengganggu konvergensi model, terutama jika jumlah sampel terbatas.

Hasil CFA menunjukkan bahwa model pengukuran awal telah memiliki kecocokan model yang baik, namun terdapat 5 butir teridentifikasi memiliki muatan faktor kurang dari 0,40. Penghapusan 5 butir yang memiliki *loading factor* $< 0,40$ terbukti meningkatkan kebersihan konstruk, konsistensi estimasi, dan reliabilitas model tanpa mengorbankan fit model secara keseluruhan. Hasil perbaikan model dengan mengeluarkan 5 butir dengan muatan faktor kurang dari 0,40 terkonfirmasi dapat memperbaiki indeks kecocokan model. Terpenuhinya indeks kecocokan model seperti CFI dan TLI lebih dari 0,90, RMSEA dan SRMR kurang dari 0,08, mengindikasikan bahwa konstruk pengukuran *scientific literacy* mampu merepresentasikan data secara memadai, dan struktur faktor yang disusun sesuai dengan data empiris. Selain itu, butir-butir yang memiliki muatan faktor lebih dari 0,40 mengonfirmasi bahwa butir-butir tersebut memiliki kontribusi signifikan dalam merepresentasikan dimensi yang diukur. Temuan-temuan tersebut menunjukkan bahwa model hasil revisi yaitu instrumen pengukuran *scientific literacy* yang terdiri dari 55 butir (setelah mengeluarkan 5 butir dengan SLF $< 0,40$) dan 4 dimensi yang saling berkorelasi memiliki bukti validitas konstruk yang memuaskan dan lebih optimal untuk digunakan sebagai instrumen pengukuran literasi sains multidimensional yang valid dan reliabel. Model revisi lebih mencerminkan struktur konseptual yang solid sesuai teori literasi sains (National Research Council, 2012; OECD, 2017).

SIMPULAN

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa perbaikan pada instrumen literasi sains PISA dengan menghapus butir-butir yang memiliki muatan faktor (SLF) kurang dari 0,40 secara signifikan meningkatkan validitas dan reliabilitas instrumen. Penggunaan *Confirmatory Factor Analysis* (CFA) menunjukkan bahwa model pengukuran revisi memberikan kecocokan yang lebih baik dengan data empiris, yang tercermin dalam penurunan nilai RMSEA dan SRMR, serta meningkatnya SLF untuk semua butir yang tersisa. Selain itu, perbaikan ini juga menghasilkan instrumen yang lebih representatif dalam mengukur literasi sains secara komprehensif dan adaptif terhadap konteks lokal, tanpa mengurangi integritas teori dan kerangka kerja global yang ada. Temuan

ini menunjukkan pentingnya evaluasi dan revisi instrumen pendidikan untuk meningkatkan kualitas pengukuran dalam penelitian internasional, serta memberikan kontribusi penting bagi pengembangan instrumen evaluasi literasi sains yang lebih valid dan reliabel di tingkat global maupun lokal.

UCAPAN TERIMA KASIH

Penulis menyampaikan penghargaan dan ucapan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan kontribusi terhadap penelitian ini, antara lain Kepala Sekolah SMPN 1 Sindang Indramayu, SMPN 2 Sindang Indramayu, SMPN 4 Sindang Indramayu, SMPN 34 Kota Bekasi, dan SMP IT Al Khoiriyah Al Husna Sukabumi, serta para ahli yang berpartisipasi dalam proses *expert judgment*.

DAFTAR RUJUKAN

- Abu Bakar, M. R., Halim, L., & Mohamad Arsad, N. (2023). Factor Influencing Students' Scientific Literacy: An Exploratory Factor Analysis. *International Journal of Academic Research in Business and Social Sciences*, 13(12), 5844–5863. <https://doi.org/10.6007/ijarbss/v13-i12/20393>
- Baumgartner, H., Weijters, B., & Pieters, R. (2018). Misresponse to Survey Questions: A Conceptual Framework and Empirical Test of the Effects of Reversals, Negations, and Polar Opposite Core Concepts. *Journal of Marketing Research*, 55(6), 869–883. <https://doi.org/10.1177/0022243718811848>
- Chao, F., Dexian, Z., Tiejun, Y., & Hongliang, F. (2014). Measurement Method of the Grain Quantity Based on The Ground Pressure. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, 12(7), 5552–5558. <https://doi.org/10.11591/telkomnika.v12i7.4079>
- Cho, G., Hwang, H., Sarstedt, M., & Ringle, C. M. (2020). Cutoff criteria for overall model fit indexes in generalized structured component analysis. *Journal of Marketing Analytics*, 8(4), 189–202. <https://doi.org/10.1057/s41270-020-00089-1>
- Douglas L. Inman. (1952). Measures for Describing the Size Distribution of Sediments. *SEPM Journal of Sedimentary Research*, Vol. 22(3), 125–145. <https://doi.org/10.1306/d42694db-2b26-11d7-8648000102c1865d>
- Evans, C. L., & Napier-Munn, T. J. (2013). Estimating error in measurements of mineral grain size distribution. *Minerals Engineering*, 52, 198–203. <https://doi.org/10.1016/j.mineng.2013.09.005>
- Fensham, P. J. (2013). International Assessments of Science Learning: Their Positive and Negative Contributions to Science Education. In *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy* (pp. 1–376). © Springer Science+Business Media. <https://doi.org/10.1007/978-94-007-6668-6>
- Han, L., & Abdul Rahman, M. N. (2025). Development and Validation of the Children's Approaches to Learning Questionnaire from Parent Ratings in China. *International Journal of Early Childhood*, 0123456789. <https://doi.org/10.1007/s13158-025-00430-8>
- Helfrich, C. D., Li, Y. F., Mohr, D. C., Meterko, M., & Sales, A. E. (2007). Assessing an organizational culture instrument based on the Competing Values Framework: Exploratory and confirmatory factor analyses. *Implementation Science*, 2(1). <https://doi.org/10.1186/1748-5908-2-13>
- Inga, C. E. C., Walton, G., & Holley, E. (2021). Statistical assessment of the effects of grain-structure representation and micro-properties on the behavior of bonded block models for brittle rock damage prediction. *Sustainability (Switzerland)*, 13(14). <https://doi.org/10.3390/su13147889>
- Jomtarak, C., Suwanjan, P., Pupat, P., & Jomtarak, R. (2019). Testing Construct Validity of

- Competency Indicators of Scientific Literacy in Junior Secondary Students. *Mediterranean Journal of Social Sciences*, 10(2), 81–92. <https://doi.org/10.2478/mjss-2019-0025>
- Koskenniska, S., Seppälä, O., & Kömi, J. (2020). A study on grain growth using a novel grain size calculation tool. *Procedia Manufacturing*, 50(2019), 684–688. <https://doi.org/10.1016/j.promfg.2020.08.123>
- Lai, K., & Green, S. B. (2016). The Problem with Having Two Watches: Assessment of Fit When RMSEA and CFI Disagree. *Multivariate Behavioral Research*, 51(2–3), 220–239. <https://doi.org/10.1080/00273171.2015.1134306>
- Lorenzo-Seva, U. (2025). Mild-restricted confirmatory factor analysis: a new approach to assessing population factor models. *Communications in Statistics: Simulation and Computation*, 0(0), 1–19. <https://doi.org/10.1080/03610918.2025.2453473>
- Manz, J. A., Todd, M. J., Iverson, L., Ball, S. J., Manning, L., & Topp, R. (2025). Validity and reliability testing of the Creighton Competency Evaluation Instrument 2.0© (CCEI 2.0). *Clinical Simulation in Nursing*, 103, 101736. <https://doi.org/10.1016/j.ecns.2025.101736>
- National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. The National Academies Press. <https://doi.org/10.17226/13165>
- Niemand, T., & Mai, R. (2018). Flexible cutoff values for fit indices in the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 46(6), 1148–1172. <https://doi.org/10.1007/s11747-018-0602-9>
- Nurhayati, R., Aw, S., Dwiningrum, S. I. A., Hajaroh, M., & Herwin, H. (2024). Validity and Reliability of Child-Friendly School Policy Evaluation Instruments in Primary Schools: Confirmatory Factor Analysis. *International Journal of Educational Methodology*, volume-10- (volume-10-issue-1-february-2024), 165–177. <https://doi.org/10.12973/ijem.10.1.965>
- OECD. (2017). PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving (Revised Edition). In *OECD Publishing*.
- OECD. (2019). *PISA 2018 Results : What Students Know and Can Do (Volume I): Vol. I*. OECD Publishing. <https://doi.org/10.1787/5f07c754-en>
- Pavlov, G., Maydeu-Olivares, A., & Shi, D. (2021). Using the Standardized Root Mean Squared Residual (SRMR) to Assess Exact Fit in Structural Equation Models. *Educational and Psychological Measurement*, 81(1), 110–130. <https://doi.org/10.1177/0013164420926231>
- Pavlů, J., Richterová, I., Šafránková, J., & Němeček, Z. (2006). Impact of surface properties on the dust grain charging. *Advances in Space Research*, 38(11), 2558–2563. <https://doi.org/10.1016/j.asr.2005.02.002>
- Peeters, M. J., & Harpe, S. E. (2020). Updating conceptions of validity and reliability. *Research in Social and Administrative Pharmacy*, 16(8), 1127–1130. <https://doi.org/10.1016/j.sapharm.2019.11.017>
- Premthaisong, S., & Srisawasdi, N. (2024). An effect of technology-infused active inquiry learning in primary school science on students’ conceptions of learning science. *Eurasia Journal of Mathematics, Science and Technology Education*, 20(6), 3–5. <https://doi.org/10.29333/EJMSTE/14662>
- Rattanachaithada, R., Kwangmuang, P., Vongtathum, P., Gamlunglert, R., & Srikoon, S. (2025). A confirmatory factor analysis of scientific critical thinking in secondary school. *Cogent Education*, 12(1). <https://doi.org/10.1080/2331186X.2025.2467508>
- Sadler, T. D., & Zeidler, D. L. (2009). Scientific literacy, PISA, and socioscientific discourse: Assessment for progressive aims of science education. *Journal of Research in Science Teaching*, 46(8), 909–921. <https://doi.org/10.1002/tea.20327>
- Sathyanarayana, & Mohanasundaram, T. (2024). Fit Indices in Structural Equation Modeling and Confirmatory Factor Analysis: Reporting Guidelines. *Asian Journal of Economics, Business and Accounting*, 24(7), 561–577. <https://doi.org/10.9734/ajeba/2024/v24i71430>
- Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research*,

- 99(6), 323–338. <https://doi.org/10.3200/JOER.99.6.323-338>
- Shi, D., DiStefano, C., Maydeu-Olivares, A., & Lee, T. (2022). Evaluating SEM Model Fit with Small Degrees of Freedom. *Multivariate Behavioral Research*, 57(2–3), 179–207. <https://doi.org/10.1080/00273171.2020.1868965>
- Shi, D., Lee, T., & Maydeu-Olivares, A. (2018). Understanding the Model Size Effect on SEM Fit Indices. *Educational and Psychological Measurement*, 79(2), 310–334. <https://doi.org/10.1177/0013164418783530>
- Sjøberg, S., & Jenkins, E. (2020). PISA: a political project and a research agenda. *Studies in Science Education*, 58(1), 1–14. <https://doi.org/10.1080/03057267.2020.1824473>
- Suwono, H., Maulidia, L., Saefi, M., Kusairi, S., & Yuenyong, C. (2022). The Development and Validation of an Instrument of Prospective Science Teachers' Perceptions of Scientific Literacy. *Eurasia Journal of Mathematics, Science and Technology Education*, 18(1). <https://doi.org/10.29333/EJMSTE/11505>
- Watson, J. C. (2017). Establishing evidence for internal structure using exploratory factor analysis. *Measurement and Evaluation in Counseling and Development*, 50(4), 232–238. <https://doi.org/10.1080/07481756.2017.1336931>
- Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, 51(1), 409–428. <https://doi.org/10.3758/s13428-018-1055-2>
- Yoon, K. L., Shaffer, V., & Benedict, A. (2020). Resolving ambiguity: Negative interpretation biases. In *Cognitive Biases in Health and Psychiatric Disorders: Neurophysiological Foundations*. Elsevier Inc. <https://doi.org/10.1016/B978-0-12-816660-4.00006-4>
- Zetterqvist, A., & Bach, F. (2023). Epistemic knowledge—a vital part of scientific literacy? *International Journal of Science Education*, 45(6), 484–501. <https://doi.org/10.1080/09500693.2023.2166372>
- Zhang, L., Liu, X., & Feng, H. (2023). Development and validation of an instrument for assessing scientific literacy from junior to senior high school. *Disciplinary and Interdisciplinary Science Education Research*, 5(1). <https://doi.org/10.1186/s43031-023-00093-2>

