

Uji Validitas Konstruk **Tes Wechsler Intelligence Scale For Children (WISC)**

Medianta Tarigan¹ & Fadillah²

¹Universitas Pendidikan Indonesia (UPI) Bandung

²Institut Teknologi Bandung (ITB) Bandung

Abstract

As one of the psychological instruments that successfully provides an overview of the level of intelligence of children, the Wechsler Intelligence Scale For Children (WISC) test needs to be evaluated its validity with psychometric test procedures. The Wechsler Intelligence Scale For Children (WISC) test consists of 10 subtests in which each subtest has many diverse items. This study included 315 kindergarten students with an age range of 5-6 years. The data obtained from it is then analyzed using confirmatory factor analysis to test the validity of the construct. This study showed that the model had a good fit with a p-value of 0.066 and an entire subtest fit.

Keywords: Intelligence; WISC; Confirmatory Factor Analysis; Reliability; Validity

Abstrak

Sebagai salah satu instrument psikologi yang berhasil memberikan gambaran mengenai tingkat kecerdasan intelegensi anak, Tes Wechsler Intelligence Scale For Children (WISC) perlu dilakukan evaluasi validitasnya dengan prosedur uji psikometris. Tes *Wechsler Intelligence Scale For Children* (WISC) terdiri dari 10 subtes yang mana tiap subtes memiliki banyak item yang beragam. Dalam penelitian ini diikutsertakan 315 partisipan yang merupakan siswa TK dengan rentang usia 5-6 tahun. Data yang diperoleh dari tersebut kemudian dianalisis dengan menggunakan analisis konfirmatori (*Confirmatory Factor Analysis*) guna menguji validitas konstruk. Penelitian ini menunjukkan bahwa secara keseluruhan model memiliki good fit dengan nilai p-value 0.066 dan seluruh subtes fit.

Kata Kunci : Inteligeensi; WISC; Analisis Faktor Konfirmatori; Reliabilitas; Validitas

Pengukuran kecerdasan digunakan untuk menentukan kemampuan belajar anak dengan mengidentifikasi kekuatan dan kelemahan pada aspek kognitif. Ketika digunakan bersama dengan informasi yang komprehensif terkait dengan wawancara guru dan orangtua, hasil tes kognitif dapat membantu penentuan program pengembangan diri, rencana pembelajaran ataupun intervensi individual apabila dibutuhkan.

Kecerdasan intelektual merupakan kapasitas gabungan atau kapasitas global seseorang untuk bertindak, bertujuan, berpikir rasional, untuk dapat menghadapi lingkungannya secara efektif (Wechsler, 1939). Kecerdasan intelektual seseorang sulit untuk dapat diketahui kecuali dengan cara mengukur perilaku yang dilakukan oleh individu. Pengukuran kecerdasan dilakukan untuk mengetahui tingkat kecerdasan intelektual pada setiap orang dalam menghadapi suatu masalah. Salah satu tes yang banyak digunakan untuk mengukur potensi intelektual anak adalah tes *Wechsler Intelligence Scale For Children (WISC)*, yang dikembangkan oleh David Wechsler. Selama lebih dari setengah abad, jika bicara tentang pengukuran kecerdasan maka skala Wechsler adalah alat ukur kecerdasan yang dianggap paling lengkap, dimana konsep, metode dan prosedur yang melekat pada tes tersebut telah menjadi acuan dalam proses pengembangan alat ukur kecerdasan lain pada masa berikutnya (Flanagan et al., 2000). Kepopuleran skala Wechsler telah membuat alat ukur ini dialihbahasakan, diadaptasi dan dikembangkan normanya untuk penggunaan di berbagai negara, bahasa dan budaya (Canivez et al., 2021). Selain dinilai sebagai alat ukur kecerdasan anak yang cukup terstandar pada masanya, alat ukur ini juga dinilai mudah diadministrasikan, menarik bagi anak, dapat menghasilkan skor kecerdasan yang lengkap, memiliki penormaan yang jelas untuk berbagai usia dan dinilai memiliki potensi untuk digunakan dalam setting klinis dengan manual yang mudah diikuti (LITTLE, 1960).

Namun tentu saja dalam proses pengembangan tersebut, banyak kritik telah dilayangkan pada alat ukur ini. Salah satu kritik besar adalah terkait konstruk teoritis yang mendasari pembuatan alat ukur ini dianggap terlalu lemah dan hasil penelitian yang terbatas terkait uji psikometriknya (Baldwin, 1940; Schafer, 1956). Pada awal penciptaannya, Wechsler tidak mendefinisikan sebuah konstruk teoritik kecerdasan yang khusus sebagai dasar pembuatan alat ukurnya. Sebaliknya, ia mendefinisikan kecerdasan itu sendiri lalu menggabungkan tes verbal dan tes performa, yang memang telah dibuat pada masa sebelumnya secara terpisah, lalu menggabungkan menjadi satu hingga membentuk satu tes kecerdasan (Bandura et al., 1999). Tidak adanya kaitan antara konstruk teoritik dan bentuk tes berlanjut hingga saat ini, yang membuat tidak adanya teori kecerdasan yang konsisten dalam proses pengembangan alat tersebut (Beaujean et al., 2018).

Di Indonesia sendiri, WISC menjadi salah satu alat ukur kecerdasan anak yang paling banyak digunakan oleh para ahli dan dipercaya mampu mendeskripsikan berbagai aspek kecerdasan anak (Nanik, 2000). Walaupun sebenarnya WISC juga memiliki keterbatasan di lapangan karena pengukuran ini harus dinilai oleh psikolog yang berpengalaman. Dengan jumlah psikolog yang terbatas dan penyebarannya tidak merata di seluruh wilayah maka hal ini menjadi masalah teknis tersendiri bagi negeri berkembang. Tidak seperti kepopulerannya dalam hal penggunaan, penelitian terkait WISC terbilang masih sangat terbatas. Hasil penelusuran literatur terkait uji psikometri pada WISC yang terpublikasikan dalam bentuk naskah ilmiah di Indonesia menunjukkan belum adanya penelitian terdahulu tentang hal ini. Secara psikometri, kriteria yang perlu dipenuhi untuk menentukan keberkualitasan sebuah alat ukur adalah validitas. Validitas akan menunjukkan apakah aitem-aitem yang menjadi pembentuk suatu instrumen diyakini mampu mengukur hal yang hendak diukur.

Tidak hanya validitas, sebuah alat ukur dapat diandalkan apabila memenuhi syarat reliabilitas, yaitu seberapa kuat sebuah instrumen mampu memberikan data yang ajeg walaupun diberikan pada waktu yang berbeda terhadap responden yang sama. Salah satu analisis yang dianggap mampu memberikan hasil yang lebih baik untuk melakukan pengujian validitas dan reliabilitas adalah melalui *Confirmatory Factor Analysis* (CFA) dalam *Structural Equation Modeling* (SEM). CFA hampir selalu digunakan dalam menguji struktur laten dari sebuah alat ukur. Melalui analisis CFA, dapat diketahui berapa faktor yang mendasari sebuah alat ukur serta pola hubungan antar aitem dan faktor (*factor loadings*). Banyaknya faktor akan menggambarkan banyaknya subtes pada sebuah alat ukur (Brown, 2006). Untuk itu, dalam penelitian kali ini akan dilakukan pengujian validitas konstruk dan reliabilitas terhadap WISC dengan menggunakan CFA sebagai metode analisisnya. Diharapkan dengan pengujian ini, dapat diperoleh informasi terkait validitas dan reliabilitas alat ukur WISC sehingga menjadi dasar penguat bagi para praktisi dalam menggunakannya di lapangan.

Setelah kesuksesan Alfred Binet menciptakan sebuah alat ukur inteligensi untuk anak (Stanford Binet), para ahli mengkritisi alat ukur tersebut dan menjadi tertarik untuk menciptakan alat yang lebih baik. Salah satunya adalah Wechsler. Alat yang pertama kali diciptakan olehnya adalah alat ukur inteligensi untuk dewasa, karena ia menganggap bahwa

alat ukur Binet hanya cocok untuk anak-anak. Tahun 1939, alat ukur tersebut dipublikasikan dengan nama Wechsler Bellevue Form 1 (W-B-1), hanya berselang 2 tahun setelah Binet melakukan revisi atas alatnya. W-B-1 diciptakan untuk individu berusia 10-60 tahun, rentang usia yang jauh lebih lebar dari alat ukur Binet tentunya. Karena adanya permintaan yang tinggi atas pengukuran individu di kalangan militer pada Perang Dunia II, Wechsler kemudian menciptakan tes paralel yang dinamakan Wechsler Adult Intelligence Scale (WAIS) (LITTLE, 1960).

Di tahun 1949, dengan mengadopsi dari konstruk tes pada W-B-1, ia menciptakan alat ukur kecerdasan khusus anak dengan rentang usia 5-15 tahun bernama WISC (Wechsler Intelligence Scale for Children). Dalam tes ini, terdapat dua bagian yaitu tes verbal dan tes performa dengan keseluruhan subtes berjumlah 12 bagian. Baik bagian tes verbal dan tes performa dapat menghasilkan indeks kecerdasan masing-masing, yaitu Verbal IQ dan Performance IQ, selain Full IQ. WISC versi awal ini yang kemudian banyak dipakai di Indonesia. Dalam konstruksinya, tidak ada informasi yang cukup dapat menjelaskan terkait teori yang mendasari terbentuknya dua indeks kecerdasan.

Dua puluh lima tahun kemudian, 1974, alat ini direvisi dengan nama WISC-R, dengan rentang usia menjadi lebih lebar yaitu 6-16 tahun. Di tahun 1991, Weschler melakukan revisi ketiga yang kemudian alat ini dinamakan WISC-III. Pada revisi ini, Weschler menyebutkan bahwa terdapat norma yang lebih representatif, subtes yang lebih baik dan aitem-aitem yang lebih sensitif terhadap kultur dan gender. Tidak hanya itu, ia juga menjanjikan bahwa adanya kejelasan stuktur faktor-faktor yang membentuk alat tes ini. Pada WISC-III, terdapat tambahan subtes baru yaitu Symbol search. Analisis faktor yang dilakukan menghasilkan 4 index skor yang baru yaitu Verbal Comprehension (VO), Perceptual Organization (PO), dan Freedom from Distractibility (FFD), dan sebagaimana yang terjadi pada WISC-R, ditambah satu index skor lagi yaitu Processing Speed (PS).

Tahun 2003, WISC revisi keempat dilakukan. Pada WISC-IV, selain Full IQ, empat index skor lain juga dihitung yaitu Verbal Comprehension (VCI), Working Memory (WMI), Perceptual Reasoning (PRI) dan Processing Speed (PSI). Terdapat 3 subtes yang dihilangkan pada WISC-IV yaitu Object Assembly, Picture Arrangement dan Mazes namun demikian

terdapat 5 tambahan subtes yang diikutsertakan sehingga pada WISC-IV terdapat 10 subtes wajib dan 5 subtes tambahan. Alasan perubahan ini adalah karena diharapkan revisi yang dilakukan dapat mengoptimalkan pengukuran kecerdasan yang lebih bersifat tidak dipengaruhi wawasan pengetahuan bersifat fakta yang dipengaruhi dengan kemampuan akademis, lebih mengukur kecerdasan fluid dan bebas budaya (Kezer & Arik, 2012). Pada pengujian psikometris dari model faktor WISC-IV ini ternyata terdapat masalah psikometris yang mempertanyakan validitas konstruk dari WISC-IV (Canivez, 2013).

Revisi paling terbaru dilakukan pada tahun 2014, yaitu WISC-V. Tujuan utama dari revisi adalah untuk memisahkan subtes Perceptual Reasoning (PR) menjadi Verbal Comprehension (VC) dan Fluid Reasoning (FR). Diharapkan bentuk ini lebih mendekati konsep teori kecerdasan Cattell–Horn–Carroll (CHC) (Canivez et al., 2021). Dengan demikian, pada edisi kelima ini terdapat lima indeks faktor kecerdasan yang didapat yaitu Verbal Comprehension (VC), Fluid Reasoning (FR), Visual Spatial (VS), Processing Speed (PS), Working Memory (WM), dan general intelligence (g) factor. Namun sekali lagi, revisi yang diajukan ternyata dianggap masih memiliki kelemahan terutama terkait dengan validitas konstruk dari lima indeks faktor kecerdasan yang ternyata dinilai masih bermasalah (Canivez dan Watkins, 2016). Melihat pada perkembangan alat tes ini, kritik dan evaluasi banyak ditujukan pada permasalahan validitas konstruk yang membentuk alat WISC.

Sebelumnya telah ada beberapa penelitian di Indonesia yang membahas mengenai alat tes WISC ini, salah satunya yaitu penelitian oleh Akhlis Munazilin (2013) dengan judul penelitian “Implementasi Sistem Pakar untuk Mengetahui Bakat Anak melalui Tes WISC (Wechsler Intelligence Scale For Children) Menggunakan Metode Forward Chaining”. Tujuan penelitian tersebut adalah untuk mempermudah pelaksanaan tes dan menjamin hasil yang akurat. Hasil yang diperoleh, system akurat lebih dari 80% dan memberi kesimpulan bahwa system pakar ini cukup baik dalam membantu tester tes WISC.

Penelitian lain dilakukan oleh Nanik (2000) yang berjudul “Penelusuran Karakteristik Hasil Tes Intelektual WISC pada Anak dengan Gangguan Pemusatan Perhatian dan Hiperaktivitas”. Penelitian ini menggunakan metode deskriptif-kuantitatif untuk mengetahui karakteristik aspek kecerdasan anak GPPH dengan menggunakan tes WISC. Hasil

menunjukkan bahwa pada beberapa anak dengan ADHD menghasilkan skor yang cukup rendah pada beberapa subtes WISC. Namun masih sangat jarang ditemukan penelitian di Indonesia yang membahas mengenai validitas konstruk pada alat tes WISC.

Dari uraian tersebut, maka penelitian ini bertujuan untuk mengukur validitas konstruk dan reliabilitas dari alat tes WISC.

Metode

Subjek Penelitian

Penelitian ini menggunakan sumber data berupa data primer yang berasal dari hasil pengetesan alat ukur WISC kepada anak-anak dengan rentang usia 5-6 tahun. Sebanyak 315 anak diikutsertakan berasal dari berbagai sekolah. Berikut sebaran subyek dalam penelitian ini.

Tabel 1.
Distribusi sampel berdasarkan umur

Kategori	Jumlah Partisipan (orang)	Presentase (%)
Usia		
5 th	242	77.83%
6 th	73	23.17%
Jenis Kelamin		
Laki-Laki	177	56.19%
Perempuan	138	43.81%

Instrumen Penelitian

WISC terdiri dari skala verbal yaitu *information, comprehension, arithmetic, similarities, vocabulary* dan *digit span*. Sementara skala performa terdiri dari *picture completion, picture arrangement, block design, object assembly, coding dan mazes*. Subtes coding dan mazes adalah subtes tambahan yang hanya diberikan apabila diperlukan tes pengganti. Pada penelitian ini, coding dan maze tidak diikutsertakan dalam analisis. Berikut uraian lebih lanjut terkait instrumen WISC

Tabel 2.
Instrumen WISC

Skala	Subtes	Jumlah aitem
Verbal	<i>Information</i>	30 aitem
	<i>Comprehention</i>	14 aitem
	<i>Arithmetic</i>	16 aitem
	<i>Similarities</i>	16 aitem
	<i>Vocabulary</i>	40 aitem
	<i>Digit span</i>	28 aitem
Performa	<i>Picture completion</i>	20 aitem
	<i>Picture arrangement</i>	11 aitem
	<i>Block design</i>	7 aitem
	<i>Object assembly</i>	4 aitem

10 subtes dikenakan pada tiap-tiap subjek. Hasil data kemudian diskor berdasarkan ketentuan manual pemberian skor WISC. Skala Verbal dan skala Performace, masing-masing akan menghasilkan IQ-Verbal dan IQ-Performansi, dan gabungan keduanya menjadi dasar untuk menghitung skor total atau Full IQ. Hasil skor skala yang diperoleh lalu dianalisis menggunakan teknik *Confirmatory Factor Analysis* (CFA) dengan dibantu software statistika JASP 0.13.1.0. Hasil analisis data ini kemudian digunakan untuk pembahasan lebih lanjut guna menjawab pertanyaan penelitian.

Untuk menganalisis validitas konstruk dari alat ukur WISC sehingga diketahui subtes mana saja yang mewakili konstruk inteligensi pada alat tersebut maka penelitian ini akan menggunakan analisis CFA. Dalam CFA, peneliti harus memiliki gambaran yang spesifik mengenai (a) jumlah faktor, (b) variabel yang mencerminkan suatu faktor, dan (c) faktor-faktor yang saling berkorelasi (Thompson, 2004). Metode analisis ini dapat digunakan untuk menguji validitas model dan mengukur hubungan antar faktor yang membentuknya (Zainudin et al., 2019). Dengan menggunakan CFA peneliti dapat menguji seberapa baik variabel yang diukur tersebut dapat mewakili konstruk yang terbentuk (Hair et al., 2010). Pada CFA, peneliti akan mengkonfirmasi atau menolak teori atau pengetahuan sebelumnya tentang model yang terbentuk dari faktor-faktor yang saling berhubungan. Berikut adalah tahapan yang dilakukan dalam analisis CFA (Henseler & Schuberth, 2020) :

1. Spesifikasi model, yaitu dengan melakukan spesifikasi variable laten dan melakukan spesifikasi kovarian dari variable laten
2. Identifikasi model, yaitu melakukan fiksasi skala yang digunakan pada variable laten dan minimal ada dua variable observasi pada setiap variable laten
3. Estimasi model dengan melakukan estimasi parameter dari model
4. Evaluasi model yaitu dengan melihat *Model fit test*, *Model fit indices*, Signifikansi parameter dan penilaian terhadap koefisien lain

Dalam menilai kesesuaian model evaluasi akan dilakukan dengan cara berikut;

1. Menilai kecocokan keseluruhan model atau *Overall Model Fit*, yang dapat dinilai dengan berbagai indeks *fit* seperti RMSEA, RMSR, GFI, CFI, TLI, NFI.
2. Menilai kecocokan model pengukuran atau *Measurement Model Fit* melalui uji *discriminant validity* melalui penilaian masing-masing konstruk laten, yang dinilai baik
3. apabila besar akar dari *Variance extracted* \sqrt{AVE} dilihat melebihi nilainya jika dibandingkan nilai korelasi antar variabel laten, sedangkan pengujian reliabilitas dengan pengujian *composit reliability* dan *variance extracted* dengan ketentuan;

$$\text{Composite reliability} = \frac{(\sum \text{standardized loading})^2}{(\sum \text{standardized loading})^2 + \sum \epsilon_j}$$

dimana *standardized loading* adalah besarnya nilai koefisien terhadap variable laten dan *measurement error* (ϵ_j) = 1 – *Standardized loading*². Serta melalui

$$\text{Variance extracted} = \frac{\sum \text{standardized loading}^2}{\sum \text{standardized loading}^2 + \sum \epsilon_j}$$

dimana *cut-off value* dari *construct reliability* adalah minimal 0,70 (Chin, 1998), dan nilai *cut-off* untuk *variance extracted* minimal 0,50 (Ghozali & Fuad, 2008).

Menilai kecocokan model struktural atau *Structural Model Fit*

Hasil

Analisis Deskriptif

Berikut merupakan hasil dari olah data untuk melihat sebaran perolehan skor dari 315 partisipan selama pengambilan data.

Tabel 3
Statistik Deskriptif

	<i>Valid</i>	<i>Mean</i>	<i>Median</i>	<i>Mode</i>	<i>Std. Dev</i>	<i>Variance</i>	<i>Min</i>	<i>Max</i>
<i>Information</i>	315	5.30	5	5	1.52	2.30	1	9
<i>Comprehension</i>	315	4.71	5	6	2.65	7.00	0	12
<i>Arithmetic</i>	315	3.68	4	4	1.46	2.14	0	8
<i>Similarities</i>	315	3.47	3	3	2.36	5.59	0	10
<i>Vocabulary</i>	315	8.54	8	7	4.47	20.01	0	21
<i>Digit Span</i>	315	2.84	3	3	1.63	2.64	0	8
<i>Picture Completion</i>	315	5.82	6	7	2.79	7.76	0	18
<i>Picture Arrangement</i>	315	6.20	6	6	2.75	7.54	0	14
<i>Block Design</i>	315	6.13	6	6	4.01	16.06	0	19
<i>Object Assembly</i>	315	10.45	9	4	6.21	38.60	0	25

Dari tabel 3, dapat dilihat bahwa nilai rata-rata tertinggi berada pada subtes *Object Assembly* dengan nilai rata-rata 10,45. Kemudian nilai rata-rata terendah diperoleh subtes *Digit Span* dengan nilai rata-rata 2,84. Untuk nilai variance, varian tertinggi juga diperoleh subtes *Object Assembly* dengan nilai varian 38,60. Nilai ini sangat tinggi jika dibandingkan dengan subtes lain, sehingga memberikan makna bahwa pada subtes ini persebaran nilai pada tiap responden sangatlah beragam.

Analisis CFA

Sebagai mana dijelaskan pada bagian sebelumnya, bahwa WISC terdiri dari 10 subtes dimana kelompok yang termasuk skala verbal adalah subtes *information*, *comprehension*, *arithmetic*, *similarities*, *vocabulary* dan *digit span*, sementara skala performa terdiri dari subtes *picture completion*, *picture arrangement*, *block design*, *object assembly*. Pada analisis berikut ini akan dilakukan analisis CFA untuk mengetahui validitas konstruk secara keseluruhan maupun secara parsial. Berikut hasil analisis yang didapatkan :

Overall model WISC

Hasil perhitungan awal untuk overall model dengan model satu faktor (unidimensional) tidak fit, karena didapatkan Chi-Square = 75.315, df = 35, P-value = 0.00008 RMSEA = 0.06. Setelah dilakukan modifikasi terhadap model dengan cara membebaskan atau memperbolehkan kesalahan pada setiap item saling berkorelasi satu dengan lainnya, maka didapatkanlah model fit dengan $P > 0,05$ (tidak signifikan). Dengan demikian model *1st order CFA* dapat diterima. Memberikan arti bahwa seluruh parameter benar-benar cocok untuk menguji variable laten WISC. Setelah dilakukan modifikasi terhadap model, dimana kesalahan pengukuran pada beberapa item dibebaskan berkorelasi satu sama lainnya, maka diperoleh model fit seperti pada tabel 4 dibawah ini:

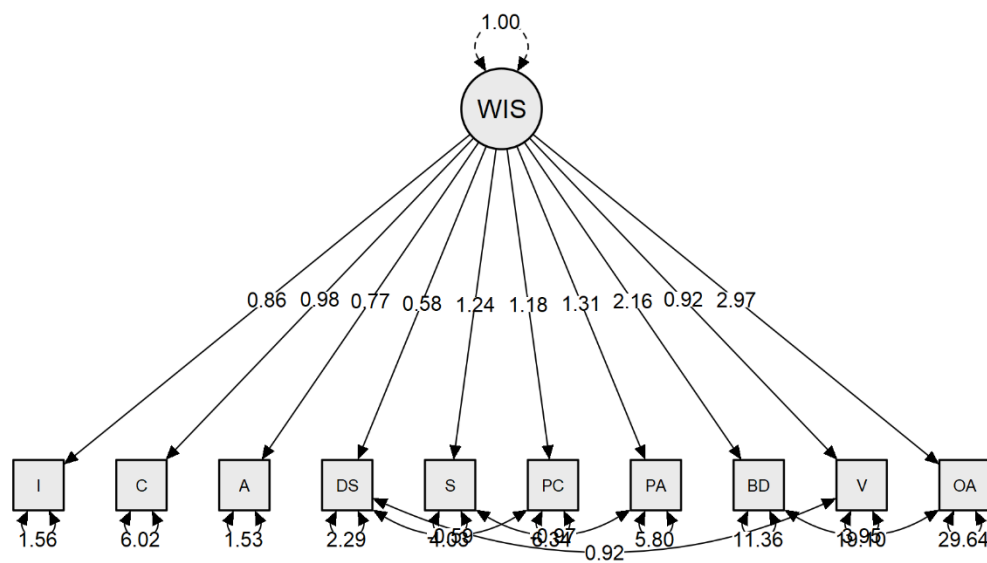
Tabel 4.
Uji Kecocokan Model Secara Keseluruhan

GOF	Tingkat Kecocokan yang dapat diterima	Indeks Model	Keterangan
Chi-square	Semakin kecil semakin baik	43.576	Baik
	p-value ≥ 0.05 model fit	0.066	Model fit
GFI	GFI ≥ 0.90 good fit	0.974	Good Fit
	$0.80 \leq \text{GFI} \leq 0.90$ marginal fit		
RMSR	RMSR ≤ 0.05 good fit	0.04	Good Fit
RMSEA	RMSEA ≤ 0.05 good fit	0.036	Good Fit
NFI	NFI $\geq 0,90$ good fit	0.897	Marginal Fit
	$0,80 \leq \text{NFI} \leq 0,90$ marginal fit		
NNFI	NNFI ≥ 0.90 good fit	0.952	Good Fit
	$0.80 \leq \text{NNFI} \leq 0.90$ marginal fit		
CFI	CFI ≥ 0.90 good fit	0.967	Good Fit
RFI	RFI $\geq 0,90$ good fit	0.85	Marginal Fit
	$0,80 \leq \text{RFI} \leq 0,90$ marginal fit		
IFI	IFI $\geq 0,90$ good fit	0.968	Good Fit
	$0,80 \leq \text{IFI} \leq 0,90$ marginal fit		
CN	CN ≥ 200 baik	326.186	Baik

Berdasarkan tabel 4 dapat dilihat bahwa *Chi-square* memiliki *p-value* $> 0,05$, nilai ini memberikan makna bahwa model yang terbentuk memiliki kecocokan yang baik. *Chi-square* bersifat sangat sensitif terhadap besarnya sampel yang digunakan, oleh karena itu χ^2 perlu dilengkapi dengan alat uji lainnya. Semakin besar ukuran sampel yang digunakan maka

statistik *Chi-square* juga akan menunjukkan nilai yang semakin besar dengan nilai *p-value* yang semakin kecil.. Atas dasar hal tersebut maka perlu dilengkapi dengan nilai *Likelihood-Ratio Chi-Square* dengan ukuran kesesuaian absolut lain, yaitu *Root Means Square Error of Approximation* (RMSEA) untuk meningkatkan keakuratan hasil pengujian *overall model fit*.. Dalam hal ini digunakan *Asymptotically Distribution Free* (ADF) yang mengabaikan asumsi normalitas data, dan pengujian tidak melihat *Chi-Square* tetapi melihat uji lain yang tidak sensitif terhadap jumlah sampel, yaitu RMSEA, SRMR, GFI, CFI, TLI.

RMSEA merupakan ukuran atau indeks yang mencoba memperbaiki karakteristik statistik *Chi-square* yang cenderung menolak model jika ukuran sampel relatif besar. Besar kecilnya nilai RMSEA tidak mempengaruhi matriks kovariansi sampel maupun matriks kovariansi populasi. Karena itu, suatu model dikatakan *fit* dengan data apabila model mampu menghasilkan nilai RMSEA yang kecil mendekati nol. Hasil analisis factor secara keseluruhan dapat dilihat pada gambar berikut :



Gambar 1. Analisis Faktor Konfirmatorik untuk model keseluruhan

Selanjutnya dilakukan uji kecocokan model pengukuran secara keseluruhan untuk mengevaluasi antara subtes dengan beberapa variable laten WISC, melihat apakah subtest tersebut signifikan mengukur faktor yang hendak diukur, sekaligus menentukan apakah

subtes tersebut perlu dikeluarkan atau tidak. Oleh karena itu, perlu dilakukan pengujian hipotesis nihil tentang koefisien muatan faktor subtes. Pengujiannya dilakukan dengan melihat *z-value* bagi setiap koefisien muatan faktor. Berikut ini hasil uji kecocokan model pengukuran :

Tabel 5.
Muatan Faktor subtes WISC secara keseluruhan

Subtes	Koefisien	Standard Error	z-Value	<i>p-value</i>	Signifikansi
<i>Information</i>	0.858	0.092	9.363	0	V
<i>Comprehension</i>	0.979	0.167	5.87	<0.001	V
<i>Arithmetic</i>	0.773	0.089	8.674	0	V
<i>Digit Span</i>	0.584	0.104	5.604	<0.001	V
<i>Similarities</i>	1.242	0.149	8.342	0	V
<i>Picture Completion</i>	1.181	0.176	6.72	<0.001	V
<i>Picture Arrangement</i>	1.308	0.176	7.444	<0.001	V
<i>Block Design</i>	2.156	0.247	8.744	0	V
<i>Vocabulary</i>	0.925	0.291	3.173	0.002	V
<i>Object Assembly</i>	2.973	0.39	7.624	<0.001	V

Keterangan: tanda V = Signifikan ($t\text{-value} > 1,96$); X = Tidak Signifikan

Dari tabel 5 dapat dilihat bahwa seluruh subtes signifikan yaitu ditunjukkan dengan nilai $t\text{-value} > 1,96$. Selain dilihat dari $t\text{-value}$, signifikansi juga bisa dilihat dari nilai $p\text{-value}$ yaitu jika bernilai kurang dari 0.05 maka subtes tersebut signifikan. Hasil juga menunjukkan bahwa semua koefisien bermuatan positif. Artinya semua koefisien muatan faktor dari subtes sesuai dengan sifat dari alat tes WISC itu sendiri, yang mana semuanya bersifat *favorable* dengan demikian semua subtes tersebut tidak akan dieliminasi.

Subtes Model WISC

Analisis CFA pada setiap subtes dari WISC diperlukan untuk mengetahui validitas internal dari setiap subtes. Hal ini untuk mengukur apakah setiap item benar-benar mengukur

subtes yang sesuai atau tidak. Sehingga jika ternyata ada item yang tidak sesuai dengan subtes maka harus dikeluarkan dari model.

Hasil perhitungan awal untuk seluruh subtes dengan model satu faktor (unidimensional) tidak fit, karena didapatkan nilai $p\text{-value} < 0.05$ dan $\text{RMSEA} > 0.05$. Setelah dilakukan modifikasi terhadap model dengan cara membebaskan atau memperbolehkan kesalahan pada setiap item saling berkorelasi satu dengan lainnya, maka didapatkanlah model fit dengan $P > 0,05$ (tidak signifikan). Rekapitulasi hasil analisis faktor konfirmatori untuk kesesuaian model dari seluruh subtes WISC disajikan pada tabel berikut :

Tabel 6.
Rekapitulasi uji kecocokan model subtes WISC

Subtes	Chi Square		p-value		RMSEA	
	Nilai	Keterangan	Nilai	Keterangan	Nilai	Keterangan
Information	116.299	Baik	0.142	Model fit	0.022	Good Fit
Comprehension	73.515	Baik	0.082	Model fit	0.029	Good Fit
Arithmetic	27.607	Baik	0.068	Model fit	0.041	Good Fit
Digit Span	69.8	Baik	0.181	Model fit	0.023	Good Fit
Similarities	48.58	Baik	0.096	Model fit	0.032	Good Fit
Picture Completion	138.598	Baik	0.095	Model fit	0.024	Good Fit
Picture Arrangement	27.652	Baik	0.068	Model fit	0.041	Good Fit
Block Design	18.33	Baik	0.106	Model fit	0.041	Good Fit
Vocabulary	591.412	Baik	0.087	Model fit	0.016	Good Fit
Object Assembly	301.182	Baik	0.231	Model fit	0.014	Good Fit

Sesuai dengan apa yang sudah dikatakan sebelumnya, bahwa model dikatakan fit jika nilai $p\text{-value}$ lebih besar dari 0.05 dan akan semakin baik jika nilai Chi-Square semakin kecil. Selain itu, model dikatakan *good fit* jika nilai RMSEA lebih kecil dari 0.05. Dari rekapitulasi hasil kecocokan model setiap subtes pada tabel 6, didapatkan hasil bahwa setelah dilakukan modifikasi indeks, seluruh subtes memiliki model yang fit. Dengan demikian model dengan hanya satu faktor dapat diterima, yang berarti bahwa seluruh item pada masing-masing subtes terbukti mengukur satu hal saja yaitu subtes itu sendiri.

Setelah diketahui bahwa setiap subtes menunjukkan model yang fit, maka perlu dilihat apakah setiap item tersebut mengukur faktor yang hendak diukur secara signifikan dan

sekaligus menentukan apakah item tersebut perlu dieliminasi atau tidak. Signifikansi item pada setiap subtes diukur dengan melihat nilai t-value dan p-value pada masing-masing item. Item-item yang masuk ke analisis signifikansi ini adalah item yang sebelumnya sudah melalui *data cleaning*. Dimana item yang memiliki nilai 0 pada seluruh responden, dieliminasi di awal karena memang tidak sesuai dengan karakteristik responden yang ada. Item-item tersebut adalah item pada subtes *Information* nomor 18 - 30, *Comprehension* nomor 13, *Arithmetic* nomor 9 - 16, *Similarities* nomor 13 - 16, *Digit Span* pada 2 nomor terakhir pada masing-masing bagian, *Picture Completion* nomor 17, *Picture Arrangement* pada 3 nomor terakhir, *Vocabulary* nomor 28, 34, 40, serta pada subtes *Object Assembly* pada nomor terakhir di setiap bagian.

Berikut rekapitulasi hasil signifikansi setiap item pada masing-masing subtes :

Tabel 7.
Rekapitulasi uji signifikansi item pada subtes WISC

Subtes	Kategori	Total Item	Persentase	No. Soal
<i>Information</i>	Signifikan	8	50%	I2, I3, I4, I5, I6, I8, I11, I15
	Tidak signifikan	8	50%	I1, I7, I9, I10, I12, I13, I14, I17
<i>Comprehension</i>	Signifikan	9	69%	C1, C2, C3, C4, C6, C7, C8, C9, C10
	Tidak signifikan	4	31%	C5, C11, C12, C14
<i>Arithmetic</i>	Signifikan	7	88%	A1, A3, A4, A5, A6, A7, A8
	Tidak signifikan	1	13%	A2
<i>Digit Span</i>	Signifikan	7	50%	DF3.1, DF3.2, DF4.1, DF5.1, DF6.1, DB2.1, DB3.1
	Tidak signifikan	7	50%	DF4.2, DF5.2, DF6.2, DF7, DB2.2, DB3.2, DB4.1
<i>Similarities</i>	Signifikan	8	73%	S1, S2, S3, S4, S5, S6, S8, S12
	Tidak signifikan	3	27%	S7, S10, S11
<i>Picture Completion</i>	Signifikan	18	95%	PC1, PC2, PC3, PC4, PC5, PC6, PC7, PC8, PC9, PC10, PC11, PC12, PC13, PC14, PC15, PC16, PC18, PC19
	Tidak signifikan	1	5%	PC20
<i>Picture Arrangement</i>	Signifikan	5	63%	PAA, PAB, PAC, PAD, PA1
	Tidak signifikan	3	38%	PA2, PA3, PA4
<i>Block Design</i>	Signifikan	7	100%	BDA, BDB, BDC, BD1, BD2, BD3, BD4
	Tidak signifikan	0	0%	-
<i>Vocabulary</i>	Signifikan	29	78%	V1, V4, V7, V8, V9, V10, V11, V12, V13, V14, V15, V16, V17, V18, V20, V21, V22, V23, V24, V26, V27, V29, V30, V31, V33,

		V35, V37, V38, V39		
Object Assembly	Tidak signifikan	8	22%	V2, V3, V5, V6, V19, V25, V32, V36
	Signifikan	25	86%	B1, B2, B3, B4, B5, B6, K1, K2, K3, K4, K5, K6, M1, M2, M3, M4, M5, M6, M7, MOB1, MOB2, MOB3, MOB4, MOB5, MOB6
	Tidak signifikan	4	14%	B7, K7, M8, MOB7

Signifikansi item pada setiap subtes diukur berdasarkan nilai z - value dan p -value pada setiap item. Item dikatakan signifikan apabila nilai z - value $> 1,96$ atau nilai p -value < 0.05 . Dari tabel 7, terlihat bahwa sebagian besar item signifikan. Bahkan pada subtes Block Design seluruh item signifikan yang ditandai dengan diperolehnya nilai t -value $> 1,96$. Hal ini dikuatkan juga dengan adanya nilai p -value < 0.05 . Sedangkan pada subtes *information* dan *digit span*, terdapat 50% item yang tidak signifikan. Kemudian 7 item lain memiliki persentase item signifikan $> 50%$ sedangkan yang tidak signifikan hanya $< 50%$. Item-item yang tidak signifikan ini lebih baik dieliminasi karena item ini tidak bisa mengukur subtes yang sesuai. Sehingga bisa dikatakan bahwa sebagian besar item signifikan dalam mengukur setiap subtes yang sesuai. Rincian hasil *loading factor* pada setiap item dapat dilihat pada lampiran. Setelah diketahui validitas secara keseluruhan maupun per subtes, maka perlu dilakukan uji reliabilitas untuk mengetahui konsistensi internal dari indikator- indikator sebuah variable.

Tabel 8.
Uji Reliabilitas

Estimate	McDonald's
Posterior mean	0.673
95% CI lower bound	0.615
95% CI upper bound	0.725

Dalam menguji reliabilitas digunakan uji konsistensi dengan menggunakan *McDonald's* pada aplikasi JASP. *McDonald's* Untuk uji reliabilitas bisa dilihat bahwa nilai realibilitas untuk itemnya sebesar 0.673, yang berarti bahwa ke 10 subtes memiliki reliabilitas yang tinggi. Dikatakan bahwa reliabilitas harus diatas 0.7, namun pada penelitian ini, nilai 0.673 masih dapat diterima karena pada *95% CI upper bound* nilai reliabilitas telah mencapai 0.725.

Pembahasan

Uji properti psikometri model WISC berbahasa Indonesia yang meliputi uji validitas dan reliabilitas ini telah dilakukan dengan melibatkan 315 partisipan yang tersebar di beberapa sekolah. Nilai Chi-Square pada uji kecocokan keseluruhan model menunjukkan nilai *Chi-square* sebesar 43,576 dengan nilai p-value 0,066. Hasil ini sesuai dengan ukuran tingkat kecocokan model, nilai yang semakin kecil semakin baik ($p\text{-value} \geq 0.05$) sehingga model ini bisa dikatakan cocok. Namun perlu diingat bahwa nilai Chi-Square sangat sensitif terhadap besaran sampel. Besaran sampel mendekati 1000 dapat dipastikan model akan ditolak, padahal hasil estimasi akan dipercaya apabila sampelnya besar (Umar & Nisa, 2020). Oleh sebab itu, disarankan untuk melihat indeks model fit lain yang relatif tidak sensitif terhadap besarnya sampel, yang paling direkomendasikan adalah dengan melihat indeks RMSEA. Dengan menggunakan metode statistika analisis faktor konfirmatori (CFA) diperoleh indeks model RMSEA 0.015 yang menunjukkan Good Fit dengan tingkat yang dapat diterima adalah $RMSEA \leq 0.05$. Hal ini juga didukung dengan indeks model lainnya yang menguatkan yaitu GFI sebesar 0.974 ($GFI \geq 0.90$), RMSR sebesar 0.04 ($RMSR \leq 0.05$), NNFI 0.952 ($NNFI \geq 0.90$) dan CFI 0.967 ($CFI \geq 0.9$). Dengan melihat hasil indeks tersebut maka dapat dikatakan bahwa model unidimensional WISC yang diuji dapat diterima. Artinya 10 dimensi pada konstruk alat ukur WISC mengukur hal yang sama, yaitu kepribadian.

Apabila melihat nilai pada hasil analisis muatan factor tiap masing-masing subtes, maka dapat dilihat bahwa hampir seluruh aitem pada WISC menunjukkan nilai z-value ≥ 1.96 . Artinya, hampir keseluruhan aitem dinilai valid dan dapat diterima. Namun ada beberapa aitem yang perlu ditelaah kembali karena terindikasi memiliki z-value yang kurang dari 1.96. Aitem pada setiap subtes tersebar secara urut dari mulai tingkat kesulitan rendah hingga tingkat kesulitan tinggi. Oleh karena itu, secara menyeluruh subtes yang tidak signifikan adalah subtes yang memiliki tingkat kesulitan tinggi sehingga menyebabkan item-item ini tidak signifikan mengukur subtes. Item pertanyaan nomor 1 pada subtes Information menjadi pengecualian karena item ini merupakan item contoh yang akan dijawab langsung oleh tester di awal sebagai contoh kepada anak-anak jika mereka belum memahami maksud dari setiap soal. Maka dari itu item ini tidak signifikan. Kemudian dari 10 subtes, jika diukur secara

individual per subtes, seluruh subtes valid karena memiliki nilai chi-square yang tidak terlalu tinggi dengan $p\text{-value} > 0,05$. Selain itu nilai RMSEA juga menunjukkan hasil *good fit*. Reliabilitas juga menunjukkan nilai yang tinggi sehingga secara keseluruhan, subtes pada WISC valid dan reliabel.

Kesimpulan

Pada penelitian ini ditemukan hasil bahwa alat ukur WISC dapat dinyatakan valid secara struktur yang membangun alatnya sekaligus valid. Untuk itu, alat ukur ini dapat direkomendasikan untuk digunakan baik oleh kalangan peneliti maupun praktisi namun perlu dilakukan koreksi dalam tiap subtes dan item yang membangunnya.

Saran

Adapun untuk penelitian selanjutnya, disarankan untuk kembali menelaah aitem-aitem dari setiap subtes sehingga lebih dipahami oleh anak-anak. Selain itu, perlu juga untuk dilakukan adaptasi penyesuaian umur dan uji validitas terhadap alat ukur WISC agar lebih tepat sasaran untuk mengukur tingkat IQ anak.

Referensi

- Baldwin, A. L. (1940). Review of The measurement of adult intelligence. *The Journal of Abnormal and Social Psychology*.
- Bandura, A., Freeman, W. H., & Lightsey, R. (1999). Self-Efficacy: The Exercise of Control. *Journal of Cognitive Psychotherapy*.
- Beaujean, A. A., Benson, N. F., McGill, R. J., & Dombrowski, S. C. (2018). A misuse of IQ scores: Using the dual discrepancy/consistency model for identifying specific learning disabilities. In *Journal of Intelligence*.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. The Guilford Press.
- Canivez, G. L. (2013). Incremental criterion validity of WAIS-IV factor index scores: Relationships with WIAT-II and WIAT-III subtest and composite scores. *Psychological Assessment*.
- Canivez, G. L., Grieder, S., & Buenger, A. (2021). Construct Validity of the German Wechsler Intelligence Scale for Children–Fifth Edition: Exploratory and Confirmatory Factor Analyses of the 15 Primary and Secondary Subtests. *Assessment*.
- Chin, W. W. (1998). The Partial LEast Squares Approach to Structural Equation Modelling. In

Modern Methods for Business Research (pp. 295–336).

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler Intelligence Scales and Gf-Gc theory: A contemporary approach to interpretation*. Allyn & Bacon.

Ghozali, I., & Fuad, F. (2008). *Structural Equation Modeling*. Universitas Diponegoro.

Hair, J., Black, W., Babin, B., & Anderson, R. (2010). *Multivariate Data Analysis Seventh Ed. In Multivariate Data Analysis: A Global Perspective*.

Henseler, J., & Schuberth, F. (2020). Using confirmatory composite analysis to assess emergent variables in business research. *Journal of Business Research*.

Kezer, F., & Arik, R. S. (2012). An Examination and Comparison of the Revisions of the Wechsler Intelligence Scale for Children. *Procedia - Social and Behavioral Sciences*.

LITTLE, M. (1960). Counter-transference *. *British Journal of Medical Psychology*, 33(1).

Munazilin, A. (2013). Implementasi Sistem Pakar Untuk Mengetahui Bakat Anak Melalui Tes WISC (Wechsler Intelligence Scale For Children) Menggunakan Metode Forward Chaining. *Jurnal Ilmiah Teknologi Informasi Asia*.

Nanik. (2000). Penelusuran Karakteristik Hasil Tes Intelligensi WISC Pada Anak Dengan Gangguan Pemusatan Perhatian dan Hiperaktivitas. *Jurnal Psikologi*.

Schafer, R. (1956). Wechsler Adult Intelligence Scale (WAIS). *Journal of Consulting Psychology*.

Thompson, B. (2004). Exploratory and confirmatory factor analysis: Understanding concepts and applications. [References]. In (2004).

Umar, J., & Nisa, Y. F. (2020). Uji validitas konstruk dengan CFA dan pelaporannya. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia*, 9(2), 1–11.

Wechsler, D. (1939). *The measurement of adult intelligence*. Williams & Wilkins Co.

Zainudin, M., Subali, B., & Jailani. (2019). Construct validity of mathematical creativity instrument: First-order and second-order confirmatory factor analysis. *International Journal of Instruction*.

Submit	Review	Revisi	Diterima	Publis
22-11-2021	22-11-2021 sd 22-02-2022	22-02-2022	22-02-2022	25-02-2022

Medianta Tarigan & Fadillah

Universitas Pendidikan (UPI) Bandung & Insitut Teknologi Bandung (ITB) Bandung

Email : medianta@upi.edu, fadillah@itb.ac.id