

---

## **Analisa Item Response Theory Wonderlic Personnel Test (WPT)**

**Medianta Tarigan<sup>1</sup>, Fadillah<sup>2</sup>**

<sup>1</sup>Universitas Pendidikan Indonesia, Indonesia

<sup>2</sup>Institut Teknologi Bandung, Indonesia

medianta@upi.edu

### **Abstract**

*Wonderlic Personnel Test (WPT) is a psychology tool that measures individual cognitive abilities based on measuring the level of learning ability, understanding the instruction, and solving the problems. In this study, WPT items were tested using the Item Response Theory (IRT) method. There were 374 participating subjects and the results of the study showed 31 items are fit with the model, while 19 items were misfit. According to the IRT 2PL model analysis, mean of examinee ability was  $-0.01$  ( $SD=1.19$ ). The mean of difficulty ( $b$ ) was  $0.48$  ( $SD=2.58$ ) and mean of discriminant ( $a$ ) was  $0.62$  ( $SD=0.38$ ). WPT test is indicated to consist of items were misfit, that do not measure the same dimension. These statistical results are in line with the characteristics of WPT which are built from three abilities to measure intelligence.*

**Keywords:** *wonderlic personnel test; item response theory; item analysis*

### **Abstrak**

Wonderlic Personnel Test (WPT) merupakan alat ukur psikologi yang mengukur kemampuan kognitif berdasarkan pada pengukuran tingkat kemampuan belajar, memahami instruksi, dan memecahkan masalah. Dalam penelitian ini dilakukan uji terhadap aitem WPT dengan metode Item Response Theory (IRT). Terdapat 374 subjek yang berpartisipasi dan hasil penelitian menunjukkan 31 aitem sesuai dengan model, sedangkan 19 aitem lagi tidak sesuai. Menurut analisis IRT model 2PL, rata-rata kemampuan peserta adalah  $-0.01$  ( $SD=1.19$ ). Sedangkan untuk rata-rata tingkat kesukaran ( $b$ ) sebesar  $0.48$  ( $SD=2.58$ ) dan rata-rata daya beda ( $a$ ) sebesar  $0.62$  ( $SD=0.38$ ). Tes WPT diindikasikan terdiri dari aitem yang tidak mengukur satu dimensi yang sama. Hasil statistik ini sejalan dengan karakteristik WPT yang dibangun dari tiga kemampuan untuk mengukur tingkat kecerdasan.

**Keywords:** *wonderlic personnel test; item response theory; analisis item*

## Pendahuluan

Kecerdasan sebagai suatu cerminan kemampuan yang luas dan mendalam untuk memahami lingkungan memegang peran penting bagi setiap individu. Hal ini membuat pengujian tingkat kecerdasan seseorang telah menjadi suatu kebutuhan dalam kehidupan sehari-hari. Untuk tujuan pengujian kecerdasan yang terstandarisasi, banyak alat pengukuran kecerdasan yang telah disarankan oleh para peneliti, salah satu diantaranya adalah alat ukur kecerdasan *Wonderlic Personnel Test* (WPT). WPT dapat menjadi alat yang menjanjikan dalam penilaian psikologis terkait tingkat kecerdasan (Dodrill, 1981). Disamping itu, alat ukur yang dikembangkan oleh E.F. Wonderlic pada tahun 1936 ini dirancang untuk membandingkan kemampuan kognitif secara umum individu dengan nilai normatif bagi beragam karir atau jenis pekerjaan berdasarkan pada pengukuran tingkat kemampuan belajar, memahami instruksi, dan memecahkan masalah (Kusdiyati, 2018).

Alat ukur yang baik salah satu indikatornya adalah bisa dinilai dari seberapa mampu alat ukur tersebut menguji subjek melalui soal dengan tingkat kesulitan mulai dari yang rendah hingga yang tinggi dan yang mampu membedakan antara subjek yang memiliki pemahaman (tentang yang diujikan) secara menyeluruh maupun tidak. Sederhananya, alat ukur yang baik adalah yang mampu membedakan responden yang berkemampuan tinggi maupun yang rendah. Untuk mengetahui informasi ini, sebuah alat ukur dapat dilakukan analisis melalui teknik *Item Response Theory* (IRT).

Beberapa penelitian telah memeriksa validitas WPT dengan membandingkannya dengan alat ukur yang berdasar pada teori kecerdasan saat ini. Salah satunya adalah yang dilakukan oleh (Matthews & Lassiter, 2008). Keduanya menyelidiki validitas dengan membandingkan WPT dan *Woodcock-Johnson Revised Tests*. Sedangkan untuk WPT berbahasa Indonesia, sebuah studi komparasi dilakukan oleh Kusdiyati pada tahun 2018. Kusdiyati dalam penelitiannya menggunakan WPT sebagai alat ukur pembanding bagi IST. Penelitian ini melaporkan bahwa terdapat perbedaan yang signifikan antara nilai IQ yang dihasilkan WPT dan IST. Perbedaan ini terjadi karena konstruksi tes WPT dan IST yang berbeda, dimana WPT dikonstruksi untuk mengukur *g-factor* sedangkan IST dikonstruksi untuk mengukur kemampuan spesifik (*s-factor* dan *g-factor*). Hingga saat ini, penelitian butir soal WPT berbahasa Indonesia dengan mengaplikasikan IRT sebagai metode analisis kualitas aitem masih terbatas. Oleh karena itu, penelitian ini bertujuan untuk mengetahui kualitas kelayakan aitem alat ukur kecerdasan WPT berbahasa Indonesia melalui teknik analisis butir soal IRT. Disamping itu, penelitian ini pun diharapkan mampu memberi manfaat berupa informasi terkini mengenai kualitas aitem WPT dan menjadi acuan bagi penelitian yang memiliki kaitan dengan WPT selanjutnya.

### *Wonderlic Personnel Test*

*Wonderlic Personnel Test* (WPT) adalah tes kemampuan kecerdasan yang dikembangkan oleh Eldon F. Wonderlic, seorang mahasiswa pasca sarjana di Universitas Northwestern pada tahun 1937. WPT yang dikembangkan oleh Wonderlic merupakan tes kemampuan kecerdasan berukuran singkat yang pertama di dunia dan dalam perkembangannya tes ini dipergunakan oleh Angkatan Laut Amerika Serikat untuk pemilihan dan pelatihan pilot selama Perang Dunia II (Chamorro-Premuzic, Dissou, Furnham, & Bales, 2010).

Wonderlic mengembangkan WPT dalam lima jenis format, yaitu A, B, D, E, dan F, dimana masing-masing terdiri dari 50 aitem soal tes dan diujikan selama 12 menit. Bentuk D, E, dan F dikembangkan dari alat ukur kemampuan mental *Otis Self-Administering Test*, sedangkan form A dan B dikembangkan secara mandiri oleh Wonderlic (Kazmier & Browne, 1959; Hicks, Harrison, & Engle, 2015). Adapun untuk versi terbaru dari WPT dirilis pada Januari 2007 dengan nama *Wonderlic Contemporary Cognitive Ability Test* berisi pertanyaan yang lebih sesuai dengan abad ke-21 ini. Di Indonesia sendiri, WPT yang banyak digunakan adalah WPT format A.

WPT terdiri dari pertanyaan yang mewakili beberapa jenis tes, diantaranya: analogi, analisis figur geometris, aritmatika, *direction following*, *disarranged sentences*, *judgement*, logika, pencocokan peribahasa,

kesamaan (*similarities*), dan definisi kata (Kazmier & Browne, 1959). Berdasarkan komponen penyusunnya diketahui bahwa WPT dikembangkan untuk mengukur kemampuan kognitif umum di bidang matematika, kosakata, dan penalaran yang selama ini telah digunakan secara luas sebagai tes kejuruan dan kecerdasan. WPT banyak digunakan untuk mengukur potensi pekerjaan, potensi pendidikan, dan potensi pelatihan seorang pelamar. Selain itu, durasi tes yang dibuat sedemikian rupa sehingga hanya sekitar dua hingga lima persen dari rata-rata kelompok menyelesaikan tes di batas waktu dua belas menit, menjadikan WPT sebagai tes kejuruan yang dapat dilangsungkan dengan singkat dan sederhana.

Setelah pertama kali ditemukan oleh Wonderlic lebih dari 70 tahun lalu, telah banyak penelitian yang dilakukan terhadap WPT. Pengembangan pada alat ukur ini pun sudah banyak dilakukan. Penelitian WPT terakhir kali dilakukan oleh T. Matthews dan Kerry Lassiter pada tahun 2007. Matthews dan Lassiter melaporkan bahwa WPT berkorelasi kuat positif secara signifikan dengan *overall intellectual functioning* tetapi tidak menunjukkan bukti validitas baik konvergen ataupun divergen terhadap dua domain kemampuan kognitif, yaitu: *fluid* dan *crystallized intelligence*. Ini mengindikasikan validitas yang kurang baik. Sedangkan untuk reliabilitasnya, berdasar pada studi yang sama diperoleh hasil bahwa pengujian reliabilitas Wonderlic menunjukkan skala reliabilitas yang cukup baik, yaitu sebesar, 0.87.

### **Item Response Theory**

*Item Response Theory* (IRT) adalah salah satu metode psikometri berdasarkan teori statistik yang terdiri dari berbagai model matematika yang pengembangannya diinisiasi oleh Frederic Lord pada 1950 dan 1960-an (Lord, 1952). Analisis ini bertujuan untuk mengevaluasi responden tanpa tergantung pada item yang sama dalam suatu tes (Zanon, Hutz, Yoo, & Hambleton, 2016). Dalam model IRT, kemampuan seseorang diestimasi dari responnya terhadap masing-masing aitem dan 'kinerja' (*performance*) setiap aitem dapat dievaluasi menggunakan parameter aitem yang digambarkan dalam suatu kurva (Yang & Kao, 2014).

IRT mensyaratkan tiga asumsi yang harus dipenuhi, yaitu unidimensi, independensi lokal, dan invariansi parameter. Asumsi unidimensi berarti hanya satu konstruk laten yang diukur oleh serangkaian aitem dalam suatu tes. Sedangkan, independensi lokal mensyaratkan bahwa respon terhadap aitem dari setiap subjek secara statistik tidak saling berhubungan dengan respon lainnya. Serta asumsi invariansi parameter adalah aitem independen dalam berbagai karakteristik sampel dalam suatu populasi (Yang & Kao, 2014).

Dengan mempertimbangkan jumlah tingkatan respon, IRT dibedakan menjadi IRT dikotomi dan politemus. Kategori dikotomi adalah bentuk tes yang memiliki jawaban benar-salah, meskipun pilihan jawaban berbentuk pilihan ganda dengan lima opsi, akan tetapi untuk model benar-salah ini tetap dikategorikan sebagai dikotomi. Sedangkan politemus adalah bentuk soal dimana para responden memungkinkan memiliki skor yang berbeda (tidak hanya benar-salah) (Ostini & Nering, 2012).

Sebagai teknik yang melibatkan pemodelan matematika, IRT melakukan estimasi parameter. Berdasarkan jumlah parameter yang diestimasi, IRT untuk data dikotomi dibagi menjadi tiga jenis model yaitu satu, dua, dan tiga parameter model logistik (1PLM, 2PLM, dan 3PLM IRT). Model logistik satu-parameter (1PLM) menduga satu jenis parameter, yaitu tingkat kesukaran aitem atau *difficulty* (*b*). Model Dua Parameter Logistik (2PLM) mengestimasi parameter *difficulty* (*b*) dan *discrimination* (*a*). Sedangkan model tiga parameter (3PLM) mengestimasi dua parameter sebelumnya (*a* dan *b*) dan parameter *guessing* (*c*) (Yang & Kao, 2014; Zanon, dkk., 2016).

Analisis IRT yang hadir sebagai solusi atas kelemahan teknik analisis aitem sebelumnya, yaitu CTT (*Classical Test Theory*). Memberikan hasil analisis lebih baik karena independensi di antara taraf kesulitan aitem dan kemampuan responden yang saling sesuai satu sama lain (Sudaryono, 2011). Sehingga, apabila diketahui tingkat kesulitan aitem maka kemampuan subjek dapat diketahui dan sebaliknya, apabila kemampuan subjek diketahui maka tingkat kesulitan aitem dapat ditentukan.

Sementara itu, alat ukur psikologis yang telah digunakan secara luas perlu dievaluasi dengan tujuan mengetahui kualitas kelayakan alat ukur tersebut. Dengan menggunakan teknik IRT akan diperoleh parameter model (tingkat kesukaran, daya beda aitem, dan *guessing*) yang menjadi indikator penilaian kualitas alat ukur tersebut.

## Metode

Penelitian analisis aitem WPT dengan teknik IRT ini menerapkan metode penelitian kuantitatif, dimana dalam penelitian ini prosedur statistik memainkan peran sentral dalam mengukur konstruk kecerdasan yang diperoleh melalui suatu prosedur tes (Fischer, Boone, Fischer, & Boone, 2014).

## Populasi, Sampel dan Teknik Pengambilan Sampel

Populasi penelitian ini adalah individu lulusan SMP hingga Strata Satu di kota Bandung. Dengan teknik pengambilan sampel *random sampling*, diperoleh sampel berjumlah 374 subjek. Jumlah ini dipilih dengan mempertimbangkan bahwa semakin besar ukuran sampel maka cenderung semakin representatif sampel tersebut terhadap populasi.

## Instrumen Penelitian

Instrumen yang diuji dalam penelitian ini adalah alat ukur kognitif Wonderlic Personnel Test (WPT) berbahasa Indonesia yang terdiri dari 50 aitem soal. Adapun jenis WPT yang digunakan adalah WPT format A. Tes ini diujikan selama 12 menit dan digunakan dalam mengukur kemampuan belajar, memahami instruksi, dan memecahkan masalah.

## Hasil dan Pembahasan

### Statistik Deskriptif

Data sampel penelitian berukuran 374 responden yang terdiri dari 24.87% pria dan 75.13% wanita. Subjek penelitian ini juga terdiri atas 59.89% lulusan SMA, 23.80% lulusan SMP, 9.89% lulusan Diploma Tiga sampai Strata Satu, dan 6.42% tidak teridentifikasi.

### Uji Asumsi IRT

IRT mensyaratkan tiga jenis asumsi yang harus dipenuhi sebelum dilakukan analisis dengan memodelkan aitem. Untuk asumsi unidimensi cukup dengan menganalisis statistik *p-value* model. Hal ini karena *p-value* menunjukkan apakah aitem sesuai dengan model IRT yang digunakan, dimana model ini dikhususkan untuk data yang bersifat unidimensional. Pada Tabel Parameter IRT 2 PL, *p-value* dinyatakan dalam kolom *sx2\_pvalue*. Adapun *p-value* yang mengindikasikan ketidaksesuaian aitem dengan model adalah aitem yang diberi tanda bintang (\*).

Uji asumsi selanjutnya yaitu asumsi independensi lokal dilakukan dengan melakukan analisis korelasi antar residual yang diestimasi dalam pemodelan IRT. Analisis ini menunjukkan hasil bahwa terdapat aitem-aitem yang berkorelasi relatif tinggi, yaitu dengan rentang nilai koefisien korelasi 0.5076 sampai dengan 0.8836. Sehingga untuk beberapa aitem tersebut asumsi independensi lokal tidak terpenuhi.

Asumsi ketiga yaitu invariansi parameter dilakukan dengan mengkorelasikan parameter IRT (*discriminant* dan *difficulty*) antara dua sub-sampel. Sub-sampel ini diperoleh dari pengelompokan sampel penelitian secara random. Hasil yang diperoleh tertera pada tabel di bawah ini.

Tabel 1 Hasil Uji Invariansi Parameter (*Difficulty* dan *Discriminant*)

<i>Difficulty</i>	b Sub_I	b Sub_II	<i>Discriminant</i>	a Sub_I	a Sub_II
b Sub_I	I	0,9555	a Sub_I	I	0,737
b Sub_II	0,9555	I	a Sub_II	0,737	I

Tabel 1 di atas menunjukkan koefisien korelasi parameter dari dua sub-sampel. Tampak bahwa setiap parameter dari kedua sub-sampel saling berkorelasi tinggi.

### Uji *Item Response Theory*

Berikut ini adalah hasil pemodelan kelayakan aitem soal alat ukur kecerdasan WPT 50 aitem dengan teknik pemodelan IRT 2 PL.

Tabel 2 Ringkasan Statistics untuk Keseluruhan Item IRT 2PLM WPT

Item	Subjek	Rata-rata Skor	Rata-rata <i>bparam</i>	Rata-rata <i>aparam</i>	Rata-rata $\theta$
50	374	16.88 (5.54)	0.48 (2.58)	0.62 (0.38)	-0.01 (1.18)

Tabel 2 menampilkan ringkasan statistik pemodelan aitem IRT dengan model 2 PL. Dengan sampel berukuran 374 subjek, dari satu set alat ukur WPT 50 aitem berbahasa Indonesia diperoleh informasi bahwa rata-rata perolehan skor tes subjek adalah 16.88. Sedangkan untuk parameter model, diperoleh hasil bahwa rata-rata tingkat kesukaran aitem sebesar 0.62 dengan standard deviasi 0.38 dan rata-rata daya beda sebesar 0.48 dan standar deviasi 2.58 serta rata-rata kemampuan peserta (*Person scoring*/ $\theta$ ) adalah -0.01 dengan standar deviasi 1.18

Tabel 3 Parameter IRT 2PLM WPT

( $n = 374, \alpha = 0.01$ )

Item	<i>aparam</i>	<i>a_se</i>	<i>bparam</i>	<i>b_se</i>	<i>sx2_fit</i>	<i>sx2_df</i>	<i>sx2_pvalue</i>
I	0,51	0,08	0,34	0,13	55,23	25,00	0,00
2	0,29	0,03	-5,98	0,16	24,10	25,00	0,51*
3	0,43	0,09	-5,90	0,92	14,83	24,00	0,93*
4	0,23	0,06	-0,80	0,34	33,45	25,00	0,12*
5	0,47	0,08	0,65	0,16	71,27	25,00	0,00
6	0,21	0,08	-4,42	1,61	20,13	25,00	0,74*
7	0,32	0,07	-1,42	0,35	38,28	25,00	0,04*
8	0,88	0,11	-0,78	0,10	122,81	24,00	0,00
9	0,33	0,08	-2,73	0,64	38,44	25,00	0,04*
10	0,16	0,02	5,97	0,26	69,25	24,00	0,00
11	0,51	0,09	-1,71	0,27	36,79	25,00	0,06*
12	1,48	0,15	-0,14	0,06	287,38	25,00	0,00
13	0,47	0,08	-1,66	0,28	40,31	25,00	0,03*
14	0,70	0,17	-3,17	0,59	21,39	25,00	0,67*
15	0,76	0,10	-0,87	0,12	64,91	25,00	0,00

Item	a <sub>param</sub>	a <sub>se</sub>	b <sub>param</sub>	b <sub>se</sub>	sx <sub>2_fit</sub>	sx <sub>2_df</sub>	sx <sub>2_pvalue</sub>
16	0,37	0,07	-0,78	0,22	49,73	24,00	0,00
17	0,36	0,08	2,00	0,44	43,23	24,00	0,01
18	0,71	0,10	0,83	0,12	88,80	25,00	0,00
19	0,40	0,07	-0,28	0,17	45,69	25,00	0,01
20	0,57	0,09	-1,00	0,17	44,11	24,00	0,01
21	0,69	0,10	1,14	0,16	73,80	24,00	0,00
22	0,32	0,07	-0,02	0,21	34,38	23,00	0,06*
23	0,54	0,09	-0,16	0,14	41,34	23,00	0,01*
24	0,93	0,12	0,48	0,09	92,31	23,00	0,00
25	0,20	0,07	0,16	0,35	29,09	23,00	0,18*
26	0,26	0,08	-1,74	0,60	29,69	22,00	0,13*
27	0,92	0,14	0,93	0,12	68,88	22,00	0,00
28	0,32	0,09	-0,09	0,25	24,52	22,00	0,32*
29	1,48	0,20	0,59	0,08	66,12	21,00	0,00
30	1,42	0,21	0,78	0,09	57,56	21,00	0,00
31	0,79	0,14	-0,28	0,13	23,55	21,00	0,32*
32	0,36	0,10	0,20	0,26	28,77	22,00	0,15*
33	0,30	0,10	1,02	0,44	21,58	22,00	0,49*
34	0,96	0,19	0,93	0,17	27,24	19,00	0,10*
35	1,49	0,64	2,67	0,64	18,09	18,00	0,45*
36	1,51	0,41	1,39	0,23	21,41	17,00	0,21*
37	1,08	0,33	1,59	0,36	32,49	17,00	0,01*
38	0,91	0,68	5,62	3,22	0,01	14,00	1,00*
39	0,15	0,12	2,58	2,55	13,10	15,00	0,59*
40	0,36	0,17	-0,03	0,45	13,77	15,00	0,54*
41	0,47	0,24	1,87	1,09	21,57	12,00	0,04*
42	0,28	0,10	5,82	1,69	11,02	10,00	0,36*
43	0,29	0,20	1,59	1,40	9,12	10,00	0,52*
44	0,29	0,20	0,25	0,85	9,10	9,00	0,43*
45	0,88	0,73	5,47	4,18	0,01	5,00	1,00*
46	0,35	0,24	1,44	1,34	3,34	5,00	0,65*
47	0,76	0,51	0,54	0,69	1,72	2,00	0,42*
48	0,84	0,78	4,72	8,20	0,00	-1,00	
49	0,58	0,54	1,52	1,87	4,77	0,00	
50	0,85	0,77	4,85	7,60	0,00	0,00	

Tabel 3 di atas menampilkan hasil pemodelan aitem IRT berupa statistik yang menerangkan kesesuaian item dengan model (*model fitting*) dan estimasi parameter setiap aitem. Statistik kesesuaian model terdiri dari *sx<sub>2\_fit</sub>* yaitu statistic chi-square. *sx<sub>2\_df</sub>* mewakili derajat kebebasan (*degree of freedom*). dan *sx<sub>2\_pvalue</sub>* menyatakan *p-value*. Sedangkan untuk estimasi parameter adalah *a<sub>param</sub>* yang mewakili *discriminant* (daya beda aitem) dan *b<sub>param</sub>* adalah estimasi *difficulty* (tingkat kesulitan aitem).

## Pembahasan

Dalam penelitian telah dilakukan analisis kelayakan aitem soal alat ukur kecerdasan WPT dengan menggunakan teknik *Item Response Theory* (IRT) dengan tujuan mengetahui kelayakan aitem WPT melalui variasi tingkat kesulitan aitem terhadap kemampuan subjek tes. Analisis kelayakan aitem soal WPT yang

terdiri dari 50 aitem dilakukan dengan berdasarkan pada sampel penelitian yang berukuran 374 subjek. Untuk satu set alat ukur WPT yang terdiri dari 50 aitem dan diujikan selama 12 menit ini, dilakukan beberapa tahapan, yaitu: uji asumsi dan uji analisis kelayakan aitem dengan metode IRT. Adapun untuk perhitungan statistik dilakukan dengan bantuan *software* jMetrik versi 4.1.1.

Sebelum melakukan analisis pemodelan IRT, terlebih dahulu perlu dipertimbangkan hasil uji asumsi IRT yang terdiri dari asumsi unidimensi, independensi lokal, dan invariansi parameter. Untuk asumsi unidimensi akan dijelaskan bersamaan dengan analisis pemodelan pada bagian selanjutnya. Adapun untuk asumsi independensi lokal, berdasarkan pada hasil pengujian dimana terdapat aitem yang saling berkorelasi kuat (koefisien korelasi lebih dari 0.5) hal ini menunjukkan bahwa respon terhadap suatu aitem dipengaruhi oleh respon terhadap aitem lainnya. Aitem-aitem ini sebagian besar adalah aitem pada bagian akhir sehingga muncul dugaan bahwa hal ini dipengaruhi tingkat kesulitan yang semakin sulit atau karena ketidakcukupan subjek tes untuk menjawab aitem di akhir tes, mengingat durasi tes WPT yang cukup singkat.

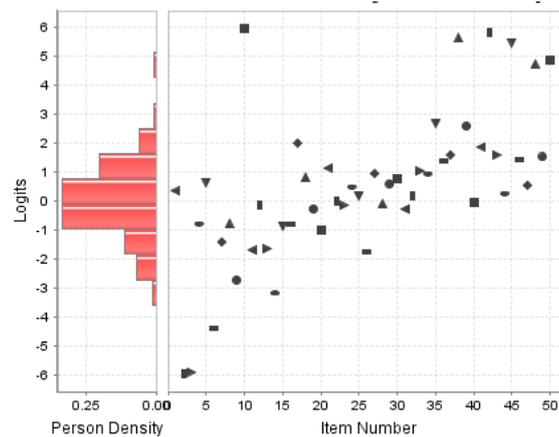
Untuk asumsi invariansi parameter, diperoleh hasil bahwa setiap estimasi parameter antara dua sub-sampel berkorelasi secara kuat. Hal ini menunjukkan asumsi dipenuhi. Dengan demikian, dapat disimpulkan bahwa aitem relatif tetap pada berbagai karakteristik sampel.

Setelah analisis dua asumsi IRT dengan beberapa catatan yang diungkapkan di atas, selanjutnya adalah analisis kelayakan aitem WPT dengan teknik IRT. Setelah dilakukan pemodelan IRT untuk skor tes WPT baik untuk IRT 1 PLM, 2PLM, dan 3PLM, disimpulkan bahwa pemodelan terbaik adalah yang dihasilkan oleh model dua parameter (2 PLM). Adapun pada bagian hasil telah ditampilkan estimasi parameter dan nilai statistik kesesuaian item dengan model IRT 2PLM yang diwakili oleh statistik  $sx2\_fi$ ,  $sx2\_df$ , dan  $sx2\_pvalue$ . Dengan hipotesis uji model:  $H_0$ : aitem mengikuti model IRT 2PLM *versus*  $H_1$ : aitem tidak mengikuti model IRT 2PL dan kriteria uji adalah *menerima*  $H_0$  jika statistik *p-value* lebih dari atau sama dengan taraf signifikansi ( $\alpha$ ). Dalam penelitian ini, taraf signifikansi yang diambil adalah  $\alpha = 0.01$ , maka  $H_0$  diterima atau dengan kata lain item sesuai dengan model IRT 2PL adalah aitem yang memiliki  $sx2\_pvalue \geq 0.01$ .

Pada tabel parameter di atas, aitem-aitem yang sesuai dengan model, yaitu yang telah diberi tanda (\*), terdapat sebanyak 31 aitem. Aitem-aitem ini dikatakan sesuai (*fit*) secara statistik dengan model IRT 2PL sementara sisanya tidak *fit* menurut model IRT 2PL. Hal ini menunjukkan bahwa menurut analisis IRT model 2PL aitem-aitem pada WPT belum cukup baik, karena hanya setengah bagian dari satu set soal yang sesuai dengan model IRT 2PL.

Disamping itu, bersamaan dengan analisis kesesuaian aitem dengan model dapat disimpulkan bahwa beberapa aitem yang tidak sesuai (*fit*) mengindikasikan bahwa set WPT terdiri dari aitem yang tidak mengukur satu dimensi yang sama. Hasil statistik ini sejalan dengan karakteristik WPT yang dibangun dari tiga kemampuan untuk mengukur tingkat kecerdasan.

Di sisi lain, dalam penelitian ini kelayakan aitem soal WPT juga dianalisis berdasarkan variasi kesulitan aitem terhadap kemampuan subjek tes. Berikut adalah *item map* yang menggambarkan hubungan antara tingkat kesulitan setiap aitem (*difficulty*) dan nilai kemampuan peserta (*Person Scoring*).



**Gambar 1.** *WPT Item Score and Person Density 2PLM Item Map*

Pada *Item map* di atas, kemampuan subjek penelitian (*Person Score*) digambarkan oleh histogram (sisi sebelah kiri) dan tingkat kesulitan aitem (*Item Difficulty*) yang diwakili oleh *scatter plot* (sisi kanan). Terlihat bahwa kemampuan subjek berkisar pada rentang nilai -3.7 sampai dengan 5.2 dan tingkat kesulitan aitem berada pada rentang -6 sampai dengan 6. Rentang nilai peserta yang lebih kecil dari rentang nilai tingkat kesulitan soal ini menunjukkan bahwa kemampuan subjek penelitian secara keseluruhan dapat diukur dengan aitem tes WPT. Dengan demikian, dari segi variasi tingkat kesulitan aitem, alat ukur WPT sudah cukup baik dalam mengukur kecerdasan subjek.

Selanjutnya, untuk aitem-aitem yang *fit* diperoleh informasi untuk aitem yang paling mudah adalah aitem nomor dua, yaitu aitem mengenai lawan kata. Sedangkan untuk aitem paling sulit adalah aitem nomor 42 yang merupakan soal geometri bangun datar. Untuk soal yang paling baik dalam membedakan tingkat kemampuan subjek tes yang ditunjukkan dengan nilai estimasi daya beda paling tinggi diantara aitem lainnya, yaitu aitem nomor 36 dimana aitem ini berisi soal cerita matematika sederhana. Serta, aitem yang paling buruk dalam membedakan kemampuan peserta adalah aitem nomor 39, yaitu aitem pemahaman kalimat.

### Limitasi

Penelitian analisis aitem WPT ini tidak terlepas dari beberapa limitasi yang membatasinya, yaitu: a) pengambilan sampel penelitian dari karakteristik subjek tes yang beragam dimana histori subjek seperti kemampuan dan hal lainnya tidak diketahui; b) sebagian besar responden tidak mengisi pada sekitar 10 (sepuluh) aitem terakhir yang disebabkan oleh keterbatasan waktu sesuai standar administrasi tes dan kemampuan subjek (rata-rata skor 16.88); dan c) terdapat beberapa aitem tidak memenuhi asumsi independensi lokal.

### Penutup

Berdasarkan analisis dan pembahasan hasil penelitian, maka dapat diperoleh kesimpulan penelitian bahwa aitem-aitem dalam alat ukur kecerdasan WPT belum cukup layak menurut model IRT 2 parameter (IRT 2 PL). Akan tetapi dalam variasi kesulitan aitem, yaitu yang dipandang dari tingkat kesukaran mencakup kemampuan subjek tes sudah cukup baik, dimana rentang kemampuan subjek tes yang berada di dalam rentang kesukaran aitem WPT.

Untuk penelitian selanjutnya, dapat disarankan untuk melakukan revisi sejumlah aitem WPT berbahasa Indonesia ini yang tidak memenuhi standar. Selain itu, dapat melakukan analisis aitem dengan mempertimbangkan WPT menjadi tiga subtes sesuai dengan tujuan pengukuran WPT yaitu mengukur tingkat kemampuan belajar, memahami instruksi, dan memecahkan masalah.

## Daftar Pustaka

- Chamorro-Premuzic, T., Dissou, G., Furnham, A., & Bales, A. (2010). *Personality traits and lay conceptions of intelligence. Personality Traits: Classifications, Effects and Changes.*
- Dodrill, C. B. (1981). An economical method for the evaluation of general intelligence in adults. *Journal of Consulting and Clinical Psychology*. <https://doi.org/10.1037/0022-006X.49.5.668>
- Fischer, H. E., Boone, W. J., Fischer, H. E., & Boone, W. J. (2014). *Quantitative Research Designs and Approaches* Quantitative Research Designs and Approaches University of Duisburg-Essen . Essen . Germany Knut Neumann Leibniz-Institute for Science and Mathematics Education ( IPN ). Kiel . Germany. (October 2015).
- Hicks, K. L., Harrison, T. L., & Engle, R. W. (2015). Wonderlic, working memory capacity, and fluid intelligence. *Intelligence*, 50, 186–195. <https://doi.org/10.1016/j.intell.2015.03.005>
- Kazmier, L. J., & Browne, C. G. (1959). Comparability of Wonderlic test forms in industrial testing. *Journal of Applied Psychology*. <https://doi.org/10.1037/h0045688>
- Kusdiyati, S. (2018). Studi Korelasi Wpt ( Wonderlic Personnel Test ) Dan Ist ( Intelligenz Struktur Test ). *Psymphatic : Jurnal Ilmiah Psikologi*, 3(1), 59–76. <https://doi.org/10.15575/psy.v3i1.2177>
- Lord, F. M. (1952). A Theory of Test Scores. *Psychometric Monograph No. 7*, (7), 84. <https://doi.org/10.1039/tf9524800166>
- Matthews, T. D., & Lassiter, K. S. (2008). What Does the Wonderlic Personnel Test Measure? *Psychological Reports*, 100(3), 707–712. <https://doi.org/10.2466/pr0.100.3.707-712>
- Ostini, R., & Nering, M. (2012). *Polytomous Item Response Theory Models. Polytomous Item Response Theory Models*. <https://doi.org/10.4135/9781412985413>
- Sudaryono. (2011). Implementasi Teori Responsi Butir ( Item Response Theory ) pada Penilaian Hasil Belajar Akhir di Sekolah. *Jurnal Pendidikan Dan Kebudayaan*, 17(16), 719–732.
- Yang, F. M., & Kao, S. T. (2014). Item response theory for measurement validity. *Shanghai Archives of Psychiatry*, 26(3), 171–177. <https://doi.org/10.3969/j.issn.1002-0829.2014.03.010>
- Zanon, C., Hutz, C. S., Yoo, H., & Hambleton, R. K. (2016). An application of item response theory to psychological test development. *Psicologia: Reflexao e Critica*, 29(1), <https://doi.org/10.1186/s41155-016-0040-x>