

Prof. Dr. Ibnu Hadjar, M.Ed.



# **Teori Klasik Dan Modern**

*dalam*

# **Pengembangan Tes**



# Tentang Penulis



## Prof. Dr. Ibnu Hadjar, M.Ed.,

Prof. Dr. Ibnu Hadjar, M.Ed., adalah guru besar dalam bidang Ilmu Evaluasi Pendidikan di Fakultas Ilmu Tarbiyah dan Keguruan (FITK) UIN Walisongo Semarang. Menekuni karirnya sebagai dosen sejak lulus dari IKIP Malang (1982). Pendidikan magister ditempuh di Queen's University, Kingston, Canada (1991). Sejak lulus Doktor bidang Penelitian dan Evaluasi Pendidikan di Universitas Negeri Jakarta (2002) juga mengajarkan mata kuliah evaluasi pendidikan pada mahasiswa, baik pada jenjang sarjana, magister, maupun doktor, termasuk di luar alamamaternya, IAIN/UIN Walisongo. Ia juga mengajarkan matakuliah terkait, seperti pengembangan instrumen/tes, metode penelitian kuantitatif, dan lain-lain. Di samping itu, ia juga banyak memberikan pelatihan pada para dosen dan guru.

Di samping mengajar, ia banyak melakukan penelitian, yang hampir seluruhnya memerlukan pengukuran, baik tes maupun kuesener. Karya ilmiahnya banyak ia sebarkan melalui buku, jurnal, dan forum ilmiah, baik di dalam maupun luar negeri. Ia juga membuat Program Aplikasi Statistik Walisongo (W-Stats), termasuk di dalamnya analisis butir tes dan non-tes.

Selama meniti karirnya, ia juga mendapatkan tugas tambahan menjabat sebagai Ketua Jurusan Pendidikan Bahasa Arab (1996-1999), Pembantu Rektor I Bidang Akademik (2003-2007), Dekan Fakultas Tarbiyah (2006-2010), Direktur Program Pascasarjana (2011-2015). Ia juga dipercaya menjadi anggota Majelis Badan Akreditasi Nasional Perguruan Tinggi/BAN-PT (2006-2012) serta anggota Dewan Guru Besar Kementerian Agama.



eureka  
media akara  
Anggota IKAPI  
No. 225/JTE/2021

0858 5343 1992  
eurekamediaaksara@gmail.com  
JL. Banjaran RT.20 RW.10  
Bojongsari - Purbalingga 53362

ISBN 978-623-487-696-3



9 786234 876963

# **TEORI KLASIK DAN MODERN DALAM PENGEMBANGAN TES**

Prof. Dr. Ibnu Hadjar, M.Ed.



**PENERBIT CV. EUREKA MEDIA AKSARA**

# **TEORI KLASIK DAN MODERN DALAM PENGEMBANGAN TES**

**Penulis : Prof. Dr. Ibnu Hadjar, M.Ed.**

**Desain Sampul : Ardyan Arya Hayuwaskita**

**Tata Letak : Herlina Sukma**

**ISBN : 978-623-487-696-3**

**Diterbitkan oleh : EUREKA MEDIA AKSARA, DESEMBER 2022  
ANGGOTA IKAPI JAWA TENGAH  
NO. 225/JTE/2021**

**Redaksi:**

Jalan Banjaran, Desa Banjaran RT 20 RW 10 Kecamatan Bojongsari  
Kabupaten Purbalingga Telp. 0858-5343-1992

Surel : eurekamediaaksara@gmail.com

Cetakan Pertama : 2022

**All right reserved**

Hak Cipta dilindungi undang-undang  
Dilarang memperbanyak atau memindahkan sebagian atau seluruh  
isi buku ini dalam bentuk apapun dan dengan cara apapun,  
termasuk memfotokopi, merekam, atau dengan teknik perekaman  
lainnya tanpa seizin tertulis dari penerbit.

## KATA PENGANTAR

*Alhamdulillah*, tiada yang patut penulis panjatkan kecuali puji syukur ke hadirat Yang Maha Mengetahui, *Al-'Alīm*, Allah SWT. Atas berkah dan rahmat-Nya penulis dapat menyelesaikan penyusunan buku yang ada di hadapan pembaca ini. Penulis berharap buku ini dapat membantu para pembaca terutama mahasiswa dan praktisi pendidikan, maupun mereka yang peduli dengan masalah-masalah pendidikan, yang pada dewasa ini memerlukan perbaikan dalam sistem evaluasi, khususnya pengukuran hasil belajar.

Dalam dunia pendidikan praktis saat ini, pembuatan tes dalam evaluasi kurang mendapat perhatian. Sekali dibuat, tes tersebut dianggap seperti meteran yang mampu memberikan informasi tentang apa yang diukur secara akurat. Tes dianggap mampu membedakan kemampuan seorang dari lainnya hanya dengan cara melihat hasil yang diperoleh testee. Hasil pengukurannya dengan tanpa pertimbangan digunakan untuk menentukan keberhasilan atau kegagalan seorang pembelajar dalam melewati proses pembejarannya. Dengan kata lain, kualitas tes tidak pernah dipermasalahkan

Kualitas tes tidak hanya ditentukan dalam proses penyusunannya maupun pelaksanaannya. Kualitas tersebut justru dapat diketahui secara akurat setelah tes tersebut digunakan. Setelah digunakan, terutama dalam ujicoba sebelum digunakan secara lebih luas, hasil tes yang berupa respon testee/peserta tes selanjutnya diolah dengan cara tertentu yang sistematis untuk mendapatkan informasi tentang kelebihan dan kekurangan dari butir-butir yang membentuknya maupun secara keseluruhan. Dari hasil pengolahan pasca pelaksanaan inilah kita dapat menggunakannya maupun untuk memperbaikinya atau membuangnya, sebelum digunakan lebih lanjut pada tes yang lain.

Teknik dan cara mengolah respon terhadap tes tersebut dikenal dengan sebutan teori tes. Dalam perkembangan teori tes, secara umum ada dua macam teori tes, klasik dan modern. Teori tes klasik telah berkembang sejak awal abad yang lalu ketika mulai banyak penggunaan tes standar, terutama di bidang pengukuran

potensi dan kompetensi psikologis dan pendidikan. Teori ini telah mendominasi praktik pengukuran dengan menggunakan tes, baik dalam penelitian maupun praktik dalam dunia pendidikan dan psikologis. Hal ini karena teori ini relatif lebih mudah secara konseptual maupun praktik penghitungannya.

Belakangan muncul teori modern yang berasal pada tahun 60-an, abad yang lalu. Sekelompok pakar bidang pengukuran psikologis yang kurang puas terhadap teori klasik, khususnya saling ketergantungan antara testee dan tes. Teori tes modern terus berkembang dan sampai saat ini banyak digunakan dalam pengukuran pendidikan dan psikologis. Hal ini karena pengolahan hasil tes dengan teori ini telah menghasilkan informasi tentang kualitas tes secara lebih akurat dan stabil sehingga hasil ukurannya lebih meyakinkan.

Pada saat ini, lembaga-lembaga pengukuran dengan tes psikologis hampir seluruhnya telah menggunakan teori ini dalam pengembangan tes yang dibuatnya, terutama untuk tes baku atau terstandar dan berkembangnya komputer untuk mengolahnya. Akan tetapi, dalam praktik dan penelitian pendidikan dan psikologi, teori ini belum banyak digunakan. Hal ini di antaranya karena konsep serta teknik analisinya rumit dan hampir tidak bisa digunakan tanpa bantuan program komputer untuk mengolahnya. Lebih dari itu, buku-buku yang membahas dan memberikan tuntunan untuk menggunakan teori ini masih sangat terbatas, terutama dalam Bahasa Indonesia. Oleh karena itu, penggunaan teori ini masih terbatas di lingkungan bidang yang secara khusus menggeluti teori tes.

Buku yang ada di hadapan pembaca ini didesain untuk membahas kedua teori tes tersebut, klasik dan modern, baik secara konseptual maupun praktis. Buku ini berusaha untuk mengungkapkan karakteristik dari masing-masing teori dan bagaimana menggunakan mereka. Diharapkan pembaca memahami kedua teori dengan baik sehingga dapat memilih mana teori yang lebih tepat secara praktis untuk kepentingan mengevaluasi dan meningkatkan kualitas tes yang dikembangkannya. Dengan demikian diharapkan tes yang digunakan dalam pengukuran

psikologis dalam bidang pendidikan dan psikologi memberikan informasi yang akurat dan terpercaya tentang objek yang diukurnya.

Dengan terselesaikannya penulisan buku ini, penulis menyampaikan terima kasih kepada pihak-pihak yang telah berkontribusi membantu proses penyelesaian buku ini, baik secara langsung maupun tidak langsung. Kepada istri tercinta, Dr. Umul Baroroh, M.Ag. serta ketiga ananda tersayang, Nabil Hajar, Nourman Hajar, dan Nadien Hajar, penulis sampaikan terima kasih pula atas pengorbanan waktu, tenaga, dan dorongannya dalam rangka penyelesaian penulisan buku ini. Rasanya tanpa bantuan dan ketulusan mereka semua tersebut, penulis tidak akan pernah mampu menyelesaikannya.

Terakhir, buku ini masih sangat memerlukan perbaikan dan penyempurnaan. Karena itu, kritik dan saran dari pembaca untuk perbaikannya sangat penulis harapkan. Walaupun menyadari masih banyak kekurangan dan keterbatasan, penulis berharap buku ini memberikan sumbangsih, betapapun kecilnya, sebagai rujukan bagi mereka yang sedang belajar untuk melakukan pengukuran, baik di bidang pendidikan maupun psikologi. *Āmīn.*

Semarang, 25 Nopember 2022  
Penulis,

Ibnu Hadjar

## DAFTAR ISI

<b>KATA PENGANTAR.....</b>	<b>iii</b>
<b>DAFTAR ISI.....</b>	<b>vi</b>
<b>DAFTAR GAMBAR.....</b>	<b>ix</b>
<b>DAFTAR TABEL .....</b>	<b>xi</b>
<b>BAB 1 TEORI TES .....</b>	<b>1</b>
A. Urgensi Teori Tes.....	1
B. Konsep Tes .....	3
C. Tujuan Tes.....	5
D. Teori Tes.....	6
E. Organisasi Penulisan.....	10
<b>BAB 2 BUTIR DAN SKOR TES .....</b>	<b>15</b>
A. Butir Tes .....	15
B. Macam-Macam Butir Tes.....	18
C. Skor Butir, Skor Tes, Dan Skor Komposit .....	26
<b>BAB 3 KONSEP STATISTIK UNTUK TES .....</b>	<b>35</b>
A. Statistik Untuk Tes .....	35
B. Statistik Skor Tunggal: Ukuran Kecenderungan Pusat	39
C. Statistik Skor Tunggal: Ukuran Penyebaran.....	42
D. Deskripsi Kinerja Individu Dengan Skor Baku-Z Dan Skor-T .....	46
E. Distribusi Normal.....	48
F. Deskripsi Hubungan Antar Dua Variabel.....	50
<b>BAB 4 KONSEP SKALA.....</b>	<b>56</b>
A. Angka Riil Dan Skala Pengukuran .....	56
B. Pengukuran Dan Skala Pengukuran.....	59
C. Tingkatan Skala Pengukuran .....	60
D. Pendekatan Penskalaan Dalam Pengembangan Test...	64
<b>BAB 5 RELIABILITAS DAN MODEL SKOR SEJATI</b>	
<b>KLASIK.....</b>	<b>68</b>
A. Teori Tes Klasik .....	68
B. Model Skor Sejati Klasik .....	69
C. Indeks Reliabilitas Dan Koefisien Reliabilitas.....	72
D. Reliabilitas Skor Komposit .....	74
E. Galat Baku Pengukuran.....	75

<b>BAB 6 PROSEDUR TAKSIRAN RELIABILITAS .....</b>	<b>78</b>
A. Prosedur Pengulangan Tes .....	78
B. Prosedur Tes Paralel .....	80
C. Prosedur Pelaksanaan Tes Tunggal .....	82
D. Reliabilitas Antar-Penilai atau Inter-Rater.....	88
E. Memaknai Koefisien .....	93
F. Faktor Yang Mempengaruhi Koefisien Reliabilitas ....	94
<b>BAB 7 PENGANTAR VALIDITAS .....</b>	<b>96</b>
A. Konsep Validitas .....	96
B. Hakikat Dan Pendekatan Validitas .....	99
C. Validita Isi.....	103
D. Validasi Kriteria .....	106
E. Validasi Konstruk .....	110
F. Faktor Yang Mempengaruhi Validitas.....	113
<b>BAB 8 ANALISIS BUTIR KLASIK.....</b>	<b>116</b>
A. Analisis Butir.....	116
B. Tingkat Kesulitan.....	117
C. Daya Beda Butir .....	120
D. Daya Pengecoh.....	128
<b>BAB 9 ANALISIS FAKTOR .....</b>	<b>135</b>
A. Pengantar Analisis Faktor .....	135
B. Konsep Dasar Analisis Faktor.....	137
C. Pendekatan Analisis Faktor .....	139
D. Langkah-Langkah Analisis Faktor .....	141
E. Asumsi Analisis Faktor .....	144
F. Melakukan Analisis Faktor: Pemuatan Faktor .....	145
G. Melakukan Analisis Faktor: Rotasi Faktor.....	153
H. Penerapan Analisis Faktor Untuk Analisis Butir Tes	155
I. Keputusan Hasil Analisis Faktor Untuk Analisis Butir Tes.....	166
<b>BAB 10 TEORI RESPON BUTIR .....</b>	<b>168</b>
A. Kritik Terhadap Teori Klasik .....	168
B. Pengantar Teori Respons Butir .....	170
C. Asumsi Utama.....	173
D. Sifat Laten Dan Kurva Karakteristik Butir/KKB .....	181
E. Ogiv Normal.....	183

F. Model Logistik.....	188
G. Model Logistik Satu-Parameter/ML1-P.....	192
H. Model Logistik Dua-Parameter/ML2-P.....	196
I. Model Logistik Tiga-Parameter/ML3-P.....	200
J. Pemilihan Model .....	205
K. Informasi Butir Dan Tes.....	207
L. Skor Testee .....	213
M. Kelebihan Dan Kelemahan TTB .....	215
N. Masa Depan Model TTB Yang Menjanjikan .....	217
<b>BAB 11 PENSKORAN HASIL TES .....</b>	<b>221</b>
A. Teori Skor Dan Penskoran .....	221
B. Penskoran Dalam Teori Klasik.....	222
C. Penskoran Formula Dan Teori Respon Butir .....	226
D. Macam-Macam Skor.....	229
E. Penafsiran Skor Tes .....	234
<b>DAFTAR PUSTAKA.....</b>	<b>240</b>
<b>GLOSARI .....</b>	<b>262</b>
<b>INDEX .....</b>	<b>271</b>
<b>TENTANG PENULIS.....</b>	<b>273</b>

## DAFTAR GAMBAR

Gambar 1	Kurva Yang Menggambarkan Penyebaran Normal ....	48
Gambar 2	Penyebaran Normal Dengan Variasi Nilai Rerata Dan Simpang Baku .....	49
Gambar 3	Nilai unit antar titik tidak harus 1, tetapi bisa sembarang angka (mis. 3, 7, 10, dst.), bahkan bisa berupa pecahan, mis. 1/7 seperti berikut ini:.....	57
Gambar 4	Garis Bilangan Real Memiliki Beberapa Sifat Penting (Crocker & Algina, 1986). ....	57
Gambar 5	Validitas Isi Dalam Capaian Belajar .....	104
Gambar 6	Jenis Validitas-Terkait Kriteria Berdasarkan Perbedaan Waktu Pengukuran Kriteria.....	108
Gambar 7	Model Analisis Faktor .....	138
Gambar 8	Pemisahan Varian .....	147
Gambar 9	Hubungan Antara Kedua Faktor Dengan Butir.....	163
Gambar 10	Hubungan Antara Kedua Faktor Yang Telah Dirotasi Dengan Butir .....	166
Gambar 11	Hubungan Antara Kemampuan Dan Respons Butir	183
Gambar 12	KKB Ogif Normal, Dengan Nilai Rerata Kemampuan = 0 Dan Simpang Baku = 1 .....	184
Gambar 13	Tiga Ogif Normal Dengan $b_1 = 0,5$ ; $b_2 = 1,0$ ; dan $b_3 = 1,5$ .....	186
Gambar 14	Tiga Ogif Normal Dengan $a_1 = 0,1$ ; $a_2 = 1,0$ ; dan $a_3 = 100$ .....	187
Gambar 15	Ogiv Dua-Parameter Untuk Butir Dengan Tingkat Kesulitan Dan Kecuraman: (1) $b_1 = -1$ ; $a_1 = 0,5$ ; (2) $b_2 = -1$ ; $a_2 = 1$ ; (3). $b_3 = 0$ ; $a_3 = 1$ ; (4). $b_4 = 2$ ; $a_4 = 1$ ; (5). $b_5 = 2,5$ ; $a_5 = 2$ ...	198
Gambar 16	Ogiv Tiga-Parameter Untuk Butir Dengan Tingkat Kesulitan, Kecuraman, Dan Tebakan Semu: (1) $b_1 = 1,5$ ;	

$a_1 = 1; c_1 = 0,5$ . (2)  $b_2 = 1; a_2 = 1; c_2 = 0,33$ . (3).  $b_3 = 1; a_3 = 1; c_3 = 0,25$ . (4).  $b_4 = 1; a_4 = ,5; c_3 = 0,2$ . (5).  $b_5 = 2; a_5 = 1; c_5 = 0,0$ .....203

Gambar 17 Kurva Karakteristik Butir dan Fungsi Informasi Butir  
dari 3 Butir Dengan Tingkat Kesulitan  $b_1 = -2, b_2 = 0,$   
dan  $b_3 = 2$  Untuk Tingkat Kemampuan Testee ( $\theta$ ) Yang  
Berbeda.....209

Gambar 18 Fungsi Informasi Tes Untuk Berbagai Tingkat  
Kemampuan Testee.....212

## DAFTAR TABEL

Tabel 1	Nilai komposit hasil belajar siswa Kelas VIII SMPN 3 (kelas Unggulan).....	29
Tabel 2	Nilai komposit hasil belajar siswa Kelas VIII SMPN 3 (kelas Unggulan).....	31
Tabel 3	Nilai komposit MIMPS siswa Kelas VIII SMPN 3 (kelas Unggulan).....	34
Tabel 4	Skor hasil tes IPA (N = 50) .....	38
Tabel 5	Persiapan penghitungan koefisien korelasi dengan menggunakan skor mentah. ....	52
Tabel 6	Data dan hasil pengitungan persiapan.....	54
Tabel 7	Hasil Penilaian Karya Tulis Ilmiah Oleh 2 Penilai.....	90
Tabel 8	Hasil Penilaian Dan Nilai Kesepatan.....	91
Tabel 9	Pendekatan validitas Tes .....	101
Tabel 10	Penafsiran Kualitas Butir Berdasarkan Tingkat Kesulitan.....	119
Tabel 11	Penafsiran indeks diskriminasi/daya beda butir .....	124
Tabel 12	Pola Jawaban 10 Testee Terhadap 3 Dari 50 Butir Tes .	124
Tabel 13	Persiapan Penghitungan Koefisien Korelasi $r_{pbis}$ .....	127
Tabel 14	Proporsi/Indeks Pengecoh Butir Tes (n = 100) .....	130
Tabel 15	Nilai Indek Pengecoh Untuk Masing-Masing Opsi Butir 1 .....	134
Tabel 16	Matriks Kovarian Antar Variabel.....	148
Tabel 17	Muatan Faktor Minimal Untuk Ukuran Sampel Yang Diperlukan Pada Taraf Signifikansi 5%.....	151
Tabel 18	Skor Yang Diperoleh Testee Untuk Setiap Butir.....	156
Tabel 19	Tafsiran Verbal Nilai Parameter Diskriminasi Butir ....	191

Tabel 20 Nilai Informasi Butir Dengan $b_i = 1$ Untuk Testee Dengan Nilai $\theta$ Tertentu .....	208
Tabel 21 Penghitungan Skor Formula Untuk 3 Testee Dengan Skor Mentah Yang Sama, Tetapi Berbeda Dalam Perilaku Menebak Pada 25 Butir Yang Memiliki 4 Pilihan Jawaban.....	224
Tabel 22 Ilustrasi Probabilitas Respons Butir Yang Benar, Besaran Skor Dengan Jumlah Jawaban Benar, Formula Skor Sejati .....	227

# BAB

# 1

# TEORI TES

## A. Urgensi Teori Tes

Sesuai dengan standar kompetensi yang harus dimiliki, setiap pendidik dituntut mampu melaksanakan penilaian terhadap hasil belajar peserta didiknya (Peraturan Menteri Pendidikan Nasional No. 16, 2006). Penilaian hasil belajar tersebut diperlukan untuk mengetahui seberapa jauh peserta didik telah mencapai tujuan pembelajaran yang telah ditetapkan. Salah satu alat yang paling penting untuk mengetahui keberhasilan tersebut adalah tes yang dirancang untuk mengukur capaian belajar siswa. Hasil pengukuran tersebut akan menjadi bahan untuk pengambilan keputusan, baik yang berkaitan dengan peserta didik maupun pembelajaran yang mereka lakukan. Karena itu, akurasi hasil penilaian belajar sangat tergantung pada kualitas tes yang digunakan untuk mengukurnya. Tes yang baik akan dapat memberikan gambaran hasil belajar dengan akurasi yang tinggi, dan sebaliknya. Karen tes harus dibuat oleh guru, maka kemampuan guru untuk membuat tes juga menentukan kualitas tes yang dibuatnya dan konsekuensinya pada akurasi hasil pengukurannya untuk dapat menggambarkan kemampuan/penguasaan siswa atas capaian belajar yang telah ditetapkan. Kualitas tes juga menjadi indikator yang penting untuk kemampuan profesional pembuatnya/guru (Arof, 2016).

Bagaimanakah kemampuan guru dalam membuat tes? Walaupun setiap guru pasti memiliki pengalaman membuat tes hasil belajar, hasil penelitian yang ada menunjukkan bahwa kemampuan mereka untuk membuat tes yang baik masih

# BAB

# 2

## BUTIR DAN SKOR TES

### A. Butir Tes

Sebuah tes pada umumnya berisi sekumpulan pertanyaan, yang secara bersamaan untuk mengukur kemampuan tertentu yang dimiliki oleh testee (peserta tes). Butir merupakan unit dasar interaksi pada tes (CARLA, 2022). Apa yang sering kita sebut sebagai pertanyaan tes lebih tepat disebut sebagai butir, karena mungkin bukan pertanyaan yang sebenarnya. Umpam balik siswa juga lebih tepat dikenal sebagai respons daripada jawaban. Soal dapat ditulis dalam berbagai format, antara lain pilihan ganda, menjodohkan, benar/salah, jawaban singkat, dan esai.

Butir tes mengacu pada pertanyaan tunggal yang ada dalam sebuah tes. Butir tes merupakan unit dasar dari tes psikologis maupun pendidikan. *Butir tes adalah tugas khusus yang diminta untuk dilakukan oleh testee atau peserta tes.* Butir tes dapat menilai satu atau lebih poin atau tujuan, dan butir sebenarnya itu sendiri dapat mengambil konstelasi yang berbeda tergantung pada konteksnya. Dalam tes bahasa, misalnya, suatu butir dapat menguji satu poin (pemahaman kosa kata yang diberikan) atau beberapa poin (kemampuan untuk memperoleh fakta dari suatu bagian dan kemudian membuat kesimpulan berdasarkan fakta tersebut). Demikian juga, tujuan yang diberikan dapat diuji dengan serangkaian butir. Misalnya, mungkin ada lima butir yang semuanya menguji satu poin tata bahasa (mis., bentuk kalimat pertanyaan). Butir dari jenis yang sama juga dapat dikelompokkan bersama untuk membentuk subtes dalam tes yang diberikan.

# BAB

# 3

## KONSEP STATISTIK UNTUK TES

### A. Statistik Untuk Tes

Analisis statistik merupakan alat yang ampuh yang digunakan lembaga pendidikan dan organisasi untuk memahami data dan memandu pengambilan keputusan. Ada berbagai jenis teknik analisis statistik yang dapat diterapkan pada berbagai data, industri, dan aplikasi. Mengetahui metode analisis statistik yang berbeda dan cara menggunakannya dapat membantu kita menjelajahi data, menemukan pola, dan menemukan tren dalam pengembangan tes.

Pengembangan dan aplikasi metode statistik terhadap permasalahan mengkonstruksi tes ujian dan menerapkan skor dari ujian ini telah berkembang pesat sejak karya perintis dari para pakar seperti Galton, Pearson, Thorndike, dan Spearman (Flanagan, 1941). Namun, jelas bagi mereka yang menangani masalah ini bahwa perkembangan ini masih jauh dari mencapai puncaknya, masih dalam periode pertumbuhan yang cepat. Kemajuan baru-baru ini terutama ditandai dengan penggantian prosedur yang lebih efisien dan ketat untuk metode empiris mentah yang muncul selama perkembangan pesat bidang ini. Prediksi keuntungan yang lebih besar dalam waktu dekat didasarkan pada relevansi untuk prosedur ini dari cabang teori statistik yang berkembang pesat yang terkait dengan analisis dan analisis faktor. Pesatnya pertumbuhan di bidang ini banyak ditunjukkan dengan tabulasi jumlah referensi yang terkait dengan topik ini termasuk, misalnya dalam jurnal Review of Educational Research.

# BAB

# 4

# KONSEP SKALA

## A. Angka Riil Dan Skala Pengukuran

Dalam sistem angka nyata/riil, angka terdiri dari dua: integer dan pecahan. Integer merupakan angka utuh seperti 0, 1, 2, 10, 999, dan seterusnya. Angka integer dapat bernilai negatif (nilai kurang dari 0) atau positif (nilai di atas 0). Angka 0 tidak positif maupun negatif. Sedangkan pecahan adalah nilai numerik yang merupakan bagian dari keseluruhan. Pecahan diperoleh dengan membagi keseluruhan menjadi beberapa bagian. Misalnya,  $\frac{1}{2}$  mewakili setengah dari bilangan bulat atau sesuatu. Pecahan dapat dinyatakan dengan angka sederhana, seringkali disebut rasio (seperti  $\frac{1}{2}$ ), desimal (seperti 0,5), persen (seperti 50%).

Sistem nomor nyata dapat direpresentasikan secara grafis dengan menggunakan garis angka nyata, satu kontinum yang dapat dibagi secara tak terhingga menjadi bagian yang lebih kecil dan lebih kecil lagi. Setiap nilai/angka dalam sistem bilangan real memiliki lokasi yang unik pada kontinum ini. (Crocker & Algina, 1986). Garis angka nyata memungkinkan kita untuk menampilkan bilangan nyata secara visual dengan menghubungkannya dengan titik-titik atau tanda lain yang unik pada sebuah garis. Panjang garis ditandai, misalnya dengan titik, yang konsisten di kedua sisi titik asal (angka 0), dan setiap tanda diberi label untuk menentukan *skala*. Bilangan real positif terletak di sebelah kanan titik asal dan bilangan real negatif terletak di sebelah kiri. Angka nol (0) tidak positif atau negatif (The Mathematic Library, 2021). Contoh garis angka nyata adalah sebagai berikut:

# BAB

# 5

## RELIABILITAS DAN MODEL SKOR SEJATI KLASIK

### A. Teori Tes Klasik

Teori tes membahas metode dan kriteria statistik untuk desain, evaluasi, dan perbandingan "tes", yaitu prosedur standar untuk mengukur konstruksi yang diminati, biasanya mencakup beberapa biasanya tes (Frey, 2020). Tes adalah prosedur untuk mengukur observasi atau konstruksi, yang mencakup beberapa butir kemudian digabungkan menjadi skor tes total. Metode dan kriteria ini didasarkan pada asumsi tentang komposisi pengukuran, sifat komponen ini, dan hubungan di antara komponen tersebut. Biasanya dibahas berkaitan dengan konstruksi psikologis seperti kemampuan, sikap atau pengetahuan, teori tes pada prinsipnya berlaku untuk pengukuran semua jenis variabel.

Ada banyak teori tes yang telah dikemukakan oleh para ahli. Akan tetapi, kerangka teori tes yang paling banyak dibahas adalah *teori tes klasik/TTK* atau *classical test theory/CTT* dan teori respons butir/TRB atau *item response theory/IRT* sebagai representasi dari teori tes "modern." Asumsi dasar dan prosedur TTK dikembangkan terutama selama paruh pertama abad ke-20. Dalam kerangka ini, metode dan kriteria yang berpengaruh dikembangkan untuk evaluasi tes secara keseluruhan, misalnya reliabilitas, dan untuk butir tes individual, misalnya, kesulitan butir dan diskriminasi butir. Sejak saat itu, beberapa batasan TTK yang serius telah diidentifikasi, seperti uji dan ketergantungan sampel dari parameter orang dan butir TTK, penskalaan parameter orang dan butir yang tidak dapat dibandingkan, interdependensi empiris antara parameter yang

# BAB

# 6

## PROSEDUR TAKSIRAN RELIABILITAS

### A. Prosedur Pengulangan Tes

Tujuan utama metode reliabilitas ini adalah untuk mengetahui tingkat stabilitas kinerja testee/responden dalam selang waktu tertentu. Selang waktu tersebut mungkin hanya beberapa menit, jam, hari, bulan, atau bahkan tahun. Untuk memperkirakan reliabilitas ini pada umumnya dilakukan dengan cara memberikan satu tes/instrumen secara berulang kepada satu kelompok individu yang sama dalam kesempatan yang terpisah. Karena itu, metode ini dimaksudkan untuk menghasilkan ukuran yang komparabel bagi testee/responden dalam pelaksanaan yang terpisah, dengan asumsi bahwa tidak ada intervensi yang berarti antar kedua waktu pengukuran. Indeks reliabilitas atau koefisien reliabilitas diperoleh dengan cara menghitung koefisien korelasi antara kedua set skor hasil tes/instrumen yang diperoleh pada tes/instrumen pertama dan pada tes/instrumen kedua (ulangan). Rumus korelasi yang biasa digunakan adalah Korelasi Pearson Product Moment, yang dalam perhitungannya melibatkan jumlah selisih masing-masing skor dari rerata distribusinya, simpang baku/standard deviasi kedua distribusi, dan banyaknya pasangan skor (Rowntree, 1981).

Varian galat berkesesuaian dengan fluktuasi acak kinerja dari satu kesempatan tes/instrumen ke kesempatan yang lain (Anastasi & Urbina, 1997). Variasi tersebut dapat terjadi sebagai akibat dari kondisi pelaksanaan tes/instrumen yang tak terkontrol, seperti perubahan cuaca, gangguan lingkungan, dan gangguan teknis, serta kondisi testee/responden, seperti lapar,

# BAB

# 7

# PENGANTAR

# VALIDITAS

## A. Konsep Validitas

Tes dan instrumen evaluasi lainnya banyak digunakan di lembaga pendidikan (Gronlund & Linn, 1990). Tes pencapaian, misalnya, dapat digunakan untuk seleksi, penempatan, diagnosis, atau sertifikasi penguasaan mataerai yang telah dipelajari. Tes bakat digunakan untuk memprediksi keberhasilan dalam kegiatan pembelajaran atau keberhasilan pekerjaan di masa depan. Penilaian pengembangan sosial-sosial dapat digunakan untuk memahami masalah belajar murid yang lebih baik atau untuk mengevaluasi efek dari program sekolah tertentu. Dalam psikologi, tes digunakan untuk mengukur kecerdasan, kreativitas, dan lain-lain. Sebagai alat ukur, penggunaan tes akan menghasilkan ukuran dalam bentuk skor, yang menggambarkan seberapa tingkat karakteristik yang dimiliki oleh testee dalam aspek yang diukur dengan tes tersebut.

Skor tes pada umumnya digunakan untuk menarik kesimpulan perilaku peserta tes atau testee dalam situasi yang berada di luar sesi pelaksanaan tes itu sendiri (Crocker & Algina, 2008). Tes masuk perguruan tinggi, misalnya, digunakan oleh perguruan tinggi yang bersangkutan menarik inferensi dari skor tes ke beberapa domain perilaku yang lebih besar, misalnya keberhasilan belajar mahasiswa, yang dapat diperlihatkan oleh pembuat tes sekarang atau di masa depan. Penggunaan skor tes yang bertanggung jawab mensyaratkan bahwa pengguna tes dapat membenarkan kesimpulan yang ditarik dengan alasan yang meyakinkan untuk menggunakan skor tes untuk tujuan

# BAB

# 8 | ANALISIS BUTIR KLASIK

## A. Analisis Butir

Dalam pengembangan tes yang menggunakan teori klasik, terutama tes standar, penyusunannya harus dilakukan dengan beberapa langkah agar bisa mendapatkan tes yang berkualitas tinggi untuk dijadikan bank soal. Setidaknya pengembangan tersebut harus melalui 4 tahapan: penyusunan butir, ujicoba butir, analisis butir, perbaikan butir. Penyusunan butir dilakukan dengan berpijak pada tujuan dan kompetensi atau perilaku yang diukur dan berpedoman pada ketentuan, sebagaimana telah dibahas sebelumnya. Jika butir-butir tes telah tersusun, tes tersebut harus diujicobakan secara empiris kepada sampel testee, yang diambil dari kelompok siswa yang akan diuji dengan tes tersebut. Hasil dari ujicoba tersebut dianalisis butir-butirnya. Analisis ini dimaksudkan untuk menemukan butir-butir yang baik dan yang tidak baik, berdasarkan kriteria yang telah ditetapkan sebelumnya. Hasil analisis butir ini, selanjutnya, dijadikan dasar untuk melakukan seleksi butir untuk digunakan dalam ujian selanjutnya dan untuk perbaikan butir-butir yang tidak baik. Dalam bab ini, pembahasan akan difokuskan pada tahapan yang ketiga, yakni analisis butir.

Analisis butir adalah proses mengumpulkan, meringkas, dan menggunakan informasi dari tanggapan testee untuk menilai kualitas pertanyaan dengan jawaban pilihan ganda (Sharma, 2021). Analisis item merupakan aspek penting dalam menjaga kualitas pertanyaan pilihan ganda yang diajukan kepada testee. Analisis ini merupakan metode statistik yang

# BAB

# 9

# ANALISIS FAKTOR

## A. Pengantar Analisis Faktor

Dalam pengembangan tes, butir-butirnya diharapkan memiliki hubungan satu sama lain karena mengukur hal yang sama. Pada umumnya sebuah tes memiliki banyak butir. Ketika butir-butir tersebut dihubungkan secara bivariat, maka sangat sulit menyimpan matriks dengan koefisien korelasi yang banyak tersebut di kepala. Misalnya, untuk mengeksplorasi hubungan antar butir dari sebuah tes dengan jumlah 30 butir, analisis dasar akan menghasilkan 435 korelasi. Hal ini perlu disederhanakan dengan teknik statistik lain.

Metode analisis faktor (*AF*) merupakan alat yang sangat berguna bagi para peneliti yang berusaha menetapkan alat ukur yang berkualitas tinggi dari konstruksi yang tidak teramat secara langsung (Tavakol & Wetzel, 2020). *AF* merupakan prosedur umum yang diterapkan dan dipromosikan secara luas untuk mengembangkan dan menyempurnakan instrumen penilaian untuk menghasilkan bukti validitas konstruk dari pengukuran tersebut. Dalam literatur, hubungan yang kuat antara validitas konstruk dan *FA* terdokumentasi dengan baik, karena metode ini memberikan bukti berdasarkan konten tes dan bukti berdasarkan struktur internal, komponen kunci validitas konstruk (Nunnally & Bernstein, 1994). Dari *FA*, bukti berdasarkan struktur internal dan berdasarkan bukti pada konten tes dapat diperiksa untuk memberi tahu kita apa yang sebenarnya diukur oleh instrumen - konsep abstrak yang diukur (yaitu, faktor, dimensi, atau konstruk) atau yang lainnya. Membangun validitas konstruk untuk interpretasi dari suatu

# BAB

# 10

## TEORI RESPON BUTIR

### A. Kritik Terhadap Teori Klasik

Teori tes klasik telah lama mendominasi analisis tes psikologis dan pendidikan, meskipun penyusunan tesnya telah dilakukan oleh para spesialis. Hal ini bisa dilihat dalam laporan penelitian yang telah terbit di jurnal-jurnal maupun buku-buku. Sedikit sekali penelitian yang melaporkan pengembangan instrumennya menggunakan teori modern. Mengapa demikian? Hal ini tidak terlepas dari kemudahannya dalam melakukan analisisnya serta penafsiran skor yang dihasilkan oleh teori tes tersebut.

Walaupun banyak digunakan dalam penyusunan dan pengembangan tes, teori ini memiliki beberapa kelemahan yang cukup mendasar. Kelemahan utama dari teori klasik adalah bahwa karakteristik testee (peserta tes) dan karakteristik tes tidak bisa dipisahkan satu sama lain, di mana masing-masing hanya bisa dipahami dan diinterpretasikan dalam konteks yang lain (Hambleton, Swaminathan, & Rogers, 1991). Keduanya saling tergantung, masing-masing tidak bisa dipahami secara terpisah dari yang lain. Ini merupakan kelemahan yang utama dari teori tes klasik.

Dalam teori klasik, karakteristik testee yang menjadi fokus kajiannya adalah kemampuan atau ability yang diukur dengan tes yang dianalisis. Kemampuan ini dinyatakan dengan skor sejati, yang didefinisikan sebagai nilai yang diharapkan dari kinerja testee yang teramat pada tes yang menjadi konsep. Secara operasional kemampuan didefinisikan sebagai proporsi jawaban testee pada butir yang benar dari seluruh butir tes

# BAB

# 11

## PENSKORAN HASIL TES

### A. Teori Skor Dan Penskoran

Dalam pendidikan dan psikologi, tes merupakan alat yang berupa serangkaian pertanyaan atau latihan untuk mengukur keterampilan, pengetahuan, kecerdasan, kapasitas, atau bakat individu atau kelompok. Hasil dari penggunaan alat ukur ini berupa angka yang disebut skor tes. *Skor tes* merupakan sebuah informasi yang berupa angka yang menyampaikan kinerja testee/peserta tes dalam suatu tes. Secara formal, skor tes dapat didefinisikan sebagai *ringkasan bukti yang terkandung dalam tanggapan testee terhadap butir tes yang terkait dengan konstruk atau konstruk yang diukur* (Thissen & Wainer, 2001).

Dalam teori tes klasik, skor seseorang/peserta dalam tes obyektif untuk setiap butir dilakukan dengan cara memberikan satu (1) poin untuk jawaban butir yang dijawab dengan benar dan nol (0) untuk jawaban yang salah. Skor tes seseorang testee diperoleh dari menjumlahkan skor dari seluruh butir soal dalam tes itu (Crocker & Algina, 1986). Aturan penilaian klasik ini dapat dirumuskan sebagai berikut:

$$X_a = \sum x_{ai}$$

di mana  $X_a$  adalah skor total tes untuk testee  $a$  dan  $x_{ai}$  adalah skor butir untuk testee  $a$  dan untuk butir  $i$ . Ketika butir diberi skor secara dikotomis, dengan nilai  $x = 0$  atau  $1$ , dan tidak ada *kredit* yang diberikan untuk butir yang kosong atau dihilangkan/tidak dijawab. Dengan cara penilaian ini, semua butir diberi bobot yang sama. Meskipun penilaian konvensional sederhana dan

## DAFTAR PUSTAKA

- Adegoke, B. (2013). Comparison of Item Statistics of Physics Achievement Test using Classical Test and Item Response Theory Frameworks. *Journal of Education and Practice*, 4(22), 87-96.
- Ahmad, K., Marzabadi, E. A., & Ashr, S. M. (2008). The Influence of Religiosity on Marital Satisfaction. *Journal of Social Science*, 4(2), 103-110.
- Albano, T. (2020, February 23). *Introduction to Educational and Psychological Measurement Using R*. Diambil kembali dari Theta Minus B: <https://thetaminusb.com/intro-measurement-r/>
- Al-Ghazali, I. (tt.). *Ihya' Ulumiddin. IV*. Beirut: Darul Kitab.
- Aliaga, M., & Gunderson, B. (2002). Interactive Statistics. Dalam D. Hartas, *Educational Research and Inquiry: Qualitative and Quantitative Approaches*. Upper Saddle River, N.J.: Prentice Hall.
- Allen, M., & Yen, W. (1979). *Introduction to Measurement Theory*. Monterey: Brooks/Cole.
- Anastasi, A., & Urbina, S. (1997). *Psychological Testing*. Upper Saddle River, N.J.: Prentice Hall.
- Andrade, C. (2021). Z Scores, Standard Scores, and Composite Test Scores Explained. *Indian Journal of Psychological Medicine*, 43(6), 555-557. <https://doi.org/10.1177/02537176211046525>.
- Archives of Psychology, 140 pp: 1-55. (t.thn.).
- Arikunto, S. (1989). *Prosedur Penelitian : Suatu Pendekatan Praktik*. Jakarta: Bina Aksara.
- Arof, O. I. (2016). Kontribusi kompetensi profesional dan pengetahuan tentang evaluasi terhadap kualitas tes IPA buatan guru. *Jurnal Penelitian dan Penilaian Pendidikan*, 1(1), 61-75. doi:10.22236/JPPP\_Vol1No1

- Arshad, S. (2014). Moderating Impact of Educational Level on Religiosity as an Antecedent of Ethical Behavior. *IOSR Journal of Business and Management*, 16(1), 82-88.
- Aydemir, M., & Egilmez, Ö. (2010). An important antecedent of ethical/unethical behavior: Religiosity. *Eurasian Journal of Business and Economics*, 3(6), 71-84. Diambil kembali dari <https://ejbe.org/EJBE2010Vol03No06p71AYDEMIR-EGILME Z.pdf>
- Babbie, E. (1998). *The practice of social research*. Belmont, CA: Wadsworth Pub. Co.
- Baker, F., & Kim, S. (2017). The Basics of Item Response Theory Using R. *Statistics for Social and Behavioral Sciences*. Cham, Switzerland: Springer International Publishing. doi: doi:10.1007/978-3-319-54205-8
- Balitbang Kemendikbud. (2019). Kamus Besar Bahasa Indonesia (KBBI). Jakarta: Badan Pengembangan dan Pembinaan Bahasa, Kementerian Pendidikan dan Kebudayaan. Diambil kembali dari <https://kbbi.web.id/>
- Bandalos, D., & Finney, S. (2010). Factor analysis: Exploratory and confirmatory. Dalam G. R. Hancock, & R. Mueller, *The reviewer's guide to quantitative methods in the social sciences* (hal. 93-114). New York: Routledge.
- Baroroh, U. (2012). Pengaruh Tingkat Keberagamaan Dan Fundamentalisme Keagamaan Terhadap Persepsi Tentang Kesetaraan Jender, Sikap Terhadap Nikah Sirri, Dan Respons Pada Rencana Legislasi Nikah Sirri (Penelitian pada Mahasiswa Muslim Universitas Diponegoro. *Disertasi, UIN Walisongo Semarang*.
- Baroroh, U. (2013, Nopember 10). Religiosity and fundamentalism as factors of perception about gender equality among indonesian muslim students. *Makalah, disajikan dalam The 2013 Annual Meeting of Society for the Scientific Study of Religion*. Boston, Amerika Serikat.

- Bartlett, M. (1951). The effect of standardization on a Chi-square approximation in factor analysis. *Biometrika*, 38, 337-344. <https://doi.org/10.1093/biomet/38.3-4.33>
- Batson, C. D., & Ventis, W. L. (1982). The religious experience: A social-psychological perspective. New York: Oxford University Press.
- Benoliel, J. (1985). Advancing qualitative approaches. *Western Journal of Research*, 7(2), 1-8.
- Benson, I. (1977). A comparison of three types of item analysis in test development using classical and latent trait methods. University of Florida.
- Beuchert, A., & Mendoza, J. (1979). A Monte Carlo comparison of ten item discrimination indices. *Journal of Educational Measurement*, 16, 109-118.
- Bichi, A. A. (2016). Classical Test Theory: An Introduction to Linear Modeling Approach to Test and Item Analysis. *International Journal for Social Studies.*, 2, 27-33.
- Bordens, K. S., & Abbott, B. B. (2008). Research Design and Methods: A Process Approach. New Yok: McGraw-Hill.
- Borg, W., & Gall, M. (1979). *Educational research: An introduction*. New York: London.
- Borsboom, D. (2003). Conceptual issues in psychological measurement. *Thesis, sepenuhnya internal, Universiteit van Amsterdam*.
- Brown, A. (2018). Item Response Theory approaches to test scoring and evaluating the score accuracy. Dalam P. Irwing, T. Booth, & D. Hughes, *The Wiley Handbook of Psychometric Testing* (hal. 607-638). London: John Wiley & Sons. Diambil kembali dari <https://kar.kent.ac.uk/id/eprint/44777>
- Brown, A., & Croudace, T. (2015). Scoring and estimating score precision using multidimensional IRT. Dalam S. P. Reise, *Handbook of Item Response Theory Modeling: Applications to*

- Typical Performance Assessment (a volume in the Theory Modeling: Applications to Typical Performance Assessment (a volume in the Multivariate Applications Series) (hal. 306-333). New York: Routledge/Taylor & Francis Group.*
- Brown, W. (1910). Some Experimental Results in the Correlation of Mental Abilities,. *British Journal of Psychology*, III, 296-322.
- Bryman, A., & Bell, E. (2003). *Bussiness Research Methods*. Oxford: Oxford university press.
- Bulut, O. (2015). Applying Item Response Theory Models to Entrance Examination for Graduate Studies: Practical Issues and Insights. *Journal of Measurement and Evaluation in Education and Psychology*, 6(2), 313-330. Diambil kembali dari <https://pdfs.semanticscholar.org/e2fb/07706a30959cbeb5f32f3f6f4e36fd858296.pdf>
- CARLA, Center for Advanced Research on Language Acquisition. (2022). *Test Construction*. Diambil kembali dari University of Minnesota: <https://carla.umn.edu/index.html>
- Clayton, R., & Gladden, J. (1974). The Five dimensions of religiosity: Toward demythologizing a sacred artifact. *Journal for the Scientific Study of Religion*, 13, 135-143. doi:<http://dx.doi.org/10.2307/1384375>
- Cochran, W. G. (1977). *Sampling techniques*. New York: John Wiley & Sons. Diambil kembali dari [https://renasf.fiocruz.br/sites/renasf.fiocruz.br/files/artigos/COCHRAN%2C%20W.%20Sampling%20techniques\\_compressed.pdf](https://renasf.fiocruz.br/sites/renasf.fiocruz.br/files/artigos/COCHRAN%2C%20W.%20Sampling%20techniques_compressed.pdf)
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37-46. doi:[doi:10.1177/001316446002000104](https://doi.org/10.1177/001316446002000104)
- College of Education, T. U. (2008). *Interpreting Test Scores*. Diambil kembali dari Iowa Testing Programs (ITP): [http://www.education.uiowa.edu/ITP/ITBS/itbs\\_interp\\_score.htm](http://www.education.uiowa.edu/ITP/ITBS/itbs_interp_score.htm)

- Comrey, A., & Lee, H. (1992). *A first course in factor analysis*. Hillsdale: Lawrence Erlbaum Associates.
- Congress, T. L. (2019). *Library of Congress Classification Outline*. Diambil kembali dari The Library of Congres: <https://www.loc.gov/catdir/cpsoc/lcco/>
- Cormack, D. (1991). *Cormack, D.S.* Oxford: Black Scientific.
- Courville, T. G. (2004). An Empirical Comparison of Item Response Theory and Classical Test Theory Item/Person Statistics. Unpublished Ph.D Dissertation. Unpublished Ph.D Dissertation, Texas A & M University.
- Creswell, J. W. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks, CA: Sage Publications.
- Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. Belmont: Wardworth Group.
- Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Mason, OH: Cengage Learning.
- Cronbach, L. (1960). *Essentials of psychological testing* (ke-2 ed.). Oxford: Harper.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-344.
- Cronbach, L., & Meehl, P. (1955). Construct Validity in Psychological Tests. *Psychological Bulletin*, 52, 281-302.
- Culligan, B. (2011). *Item Response Theory , Reliability and Standard Error*. Dipetik Oktober 2, 2022, dari [https://www.wordengine.jp/research/pdf/IRT\\_reliability\\_and\\_standard\\_error.pdf](https://www.wordengine.jp/research/pdf/IRT_reliability_and_standard_error.pdf).
- Dalen, V., & Deabold, V. (1979). *Understanding Educational Research*. New York: McGraw Hill Book Co.
- de Ayala, R. (2009). *The Theory and Practice of Item Response Theory*. New York, NY: The Guilford Press.

- De Ayala, R. (2009). *The theory and practice of item response theory*. New York: Guilford Press.
- Delice, A. (2010). The Sampling Issues in Quantitative Research •. *Educational Sciences: Theory & Practice*, 10(4), 2001-2018.
- DeMars, C. (2010). *Item Response Theory* . Online Edition: Oxford Academic.  
doi:<https://doi.org/10.1093/acprof:oso/9780195377033.001.0001>
- Di, D. (2020). Are Religious Women More Traditionalist? A Cross-National Examination of Gender and Religion . *Journal for the Scientific Study of Religion*, 59(4), 606-628. doi:<https://doi.org/10.1111/jssr.12683>
- Dilbeck, K. (2017). Factor Analysis: Varimax Rotation. Dalam M. Allen, *The SAGE Encyclopaedia of Communication Research Methods*.  
doi:<https://dx.doi.org/10.4135/9781483381411.n191>
- Dodge, Y. (2003). *The Oxford Dictionary of Statistical Terms*. Oxford: Oxford University Press.
- Duffy, M. (1985). Designing research the qualitative – quantitative debate. *Journal of Advanced Nursing*, 11(3), 225-232.
- Duffy, M. (1987). Methodological triangulation: a vehicle for merging quantitativeand qualitative methods. *Image*, 1-3.
- Duffy, M. (1987). Methodological triangulation: A vehicle for merging quantitativeand qualitative methods. *Image*, 9(3), 130-133.
- Duffy, M. (2029). Percobaan. *Agustus*, 20-29.
- Duke, J. T. (1998). The Dimensions of Religiosity: A Conceptual Model with an Empirical Test. Dalam *Latter-day Saint Social Life: Social Research on the LDS Church and its Members* (hal. 203-230). Provo, UT: Religious Studies Center, Brigham Young University.

- Ebel, R. (1965). *Measuring educational measurement*. Englewood Clift, N.J.: Prentice Hall.
- Echols, J. M., & Shadily, H. (2014). *Kamus Inggris-Indonesia*. Jakarta: Gramedia Pustaka Utama.
- Ellis, J. (2015). MTP2 and Partial Correlations in Monotone Higher-Order Factor Models. *Springer Proceedings in Mathematics and Statistics*. 89, hal. 261-272. Springer.
- Ellis, J. (2017). *Factor analysis and item analysis*. Dipetik Nopember 12, 2022, dari [https://www.applyingstatisticsinbehaviouralresearch.com/documenten/factor\\_analysis\\_and\\_item\\_analysis\\_versi\\_on\\_11\\_.pdf](https://www.applyingstatisticsinbehaviouralresearch.com/documenten/factor_analysis_and_item_analysis_versi_on_11_.pdf)
- El-Menouar, Y., & Stiftung, B. (2014). The Five Dimensions of Muslim Religiosity: Results of an Empirical Study . *MDA: Methods, Data, Analyses*, 8(1), 53-78. doi:10.12758/mda.2014.003
- Elrod, S. (2016). Quantitative Reasoning: The Next "Across the Curriculum" Movement. *Peer Review*, 16(3). Diambil kembali dari <https://www.aacu.org/peerreview/2014/summer/elrod>
- Fidan, A., Alagoz, R., & Karaman, N. (2021). Liberal sexual morality, religion, and attitudes toward abortion in Turkey. *Journal for the Scientific Study of Religion*, 60(2), 13–225. doi:<https://doi.org/10.1111/jssr.12755>
- Flanagan, J. (1941). Statistical Methods Related to Test Construction and Evaluation. *Review of Educational Research*, 11(1), 109-130. doi:<https://doi.org/10.3102/00346543011001109>
- Flanagan, T. (2013). (2013). The scientific method and why it matters. *C2C Journal*, 7(1), 4-6.
- Floyd, F. J., & Widaman, K. F. (1995). (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7(3), 286–299. doi:<https://doi.org/10.1037/1040-3590.7.3.286>

- Floyd, F., & Widaman, K. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7(3), 286–299. doi:<https://doi.org/10.1037/1040-3590.7.3.286>
- Frey, F. (2020). Test Theory, Classical Test Theory. Dalam J. Matthes, *International Encyclopedia of Communication Research Methods* (hal. 1-6). Wiley Online. doi:doi:10.1002/9781119011071.iemp0047
- Frost, J. (2022). *Inter-Rater Reliability: Definition, Examples & Assessing*. Diambil kembali dari Statistics By Jim: Making statistics intuitive: <https://statisticsbyjim.com/hypothesis-testing/inter-rater-reliability/>
- Glass, G. V., & Hopkins, K. D. (1984). *Statistical Methods in Education and Psychology*. Englewood Cliffs, NJ: Prentice-Hall.
- Glock, C. Y. (1972). On the Study of Religious Commitment. Dalam J. E. Faulkner, *Religion's Influence in Contemporary Society, Readings in the Sociology of Religion* (hal. 38-56). Ohio: Charles E. Merrill.
- Goetz, J., & LeCompte, M. (1984). *Ethography and Qualitative Design in Educational Research*. San Diego : Academic Press.
- Goldstein, H. (2012). Francis Galton, measurement, psychometrics and social progress. *Assessment in Education: Principles, Policy & Practice*, 19(2). doi:DOI: 10.1080/0969594X.2011.614220
- Gorsuch, R. (1983). *Factor analysis*. Hillsdale, NJ: Erlbaum.
- Gregory, R. J. (2000). *Psychological testing : history, principles, and applications*. Needham Height, MA: Allyn & Bacon.
- Gronlund, N., & Linn, R. (1990). *Measurement and Evaluation in Teaching*. New York: Macmillan.
- Grounlund, N. (1993). *How to Make Achievement Tests and Assessments*. Boston: Allyn and Bacon.
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. Dalam N. K. Denzin, & Y. S. Lincoln,

*Handbook of qualitative research* (hal. 105-117). Thousand Oaks, CA: Sage.

Hadjar. (2010). *Prasangka Keagamaan*. Semarang: Walisongo Press.

Hadjar, I. (1999). *Dasar-dasar Metodologi Penelitian Kuantitatif dalam Pendidikan*. Jakarta: Rajawali.

Hadjar, I. (2002). Pengaruh lingkungan pendidikan dan tipe kepribadian pada prasangka terhadap keolompok agama lain. Universitas Negeri Jakarta, Disertasi tidak dipublikasikan.

Hadjar, I. (2017). Religiosity, attitude towards corruption and cheating among muslim students in Indonesia. *Journal of Education and Human Development*, 6(2), 139-147. doi:DOI: 10.15640/jehd.v6n2a15

Hadjar, I. (2019). *Statistik untuk Ilmu Pendidikan, Sosial dan Humaniora*. Bantung: Rosda Karya.

Hadjar, I. (2019). *Statistik untuk ilmu pendidikan, sosial, dan humaniora*. Bandung: Remaja Rosda Karya.

Hair, J., Tatham, R., Anderson, R., & Black, W. (1998). *Multivariate data analysis*. London: Prentice-Hall.

Hambleton, R. (1994). Item Response Theory: A Broad Psychometric Framework For Measurement Advances. *Psicothema*, 6(3), 535-556. Diambil kembali dari <https://www.psicothema.com/pi? pii=946>

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: : Sage Press.

Hambleton, R., & Jones, R. (2005). An NCME Instructional Module on: Comparison of Classical Theory and Item Theory and Their Application to Test Development. *Educational Measurement Issues and Practice*, 12(3), 38-47. doi:DOI:10.1111/j.1745-3992.1993.tb00543.x

- Hambleton, R., Swaminathan, H., & Rogers, H. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Press.
- Hanzaee, K., Attar, M., & Alikhan, F. (2011). Investigating the effect of gender role attitude on the relationship between dimensions of religiosity and new product adoption intention. *World Applied Sciences Journal*, 13(6), 1527-1536.
- Hayes, S. C., Hayes, L. J., & Reese, H. W. (1988). Finding the philosophical core: A review of Stephen C. Pepper's World Hypotheses. *Journal of the Experimental Analysis of Behavior*, 50, 97-111.
- Henn, M., Weinstein, M., & Foard, N. (2006). *A short introduction to social research*. London: Sage.
- Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published. *Educational and Psychological Measurement*, 66, 393-416. doi:<https://doi.org/10.1177/0013164405282485>
- Henson, R., & Roberts, J. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, 66, 393-416. doi:<http://dx.doi.org/10.1177/0013164405282485>
- Hill, B. D. (2011). The sequential Kaiser-Meyer-Olkin procedure as an alternative for determining the number of factors in common-factor analysis: A Monte Carlo simulation. Disertasi, Oklahoma State University.
- Hill, P. C., & Hood Jr., R. W. (1999). *Measure of Religiosity*. Birmingham: Religious Education Press.
- Homaei, R., Bozorgi, Z., Ghahfarokhi, M., & Hosseinpou, S. (2016). Relationship between Optimism, Religiosity and Self-Esteem with Marital Satisfaction and Life Satisfaction. *International Education Studies*, 9(6), 53-61.

- Hu, Z., Lin, L., Wang, Y., & Li, J. (2021). The Integration of Classical Testing Theory and Item Response Theory . *Psychology*, 12, 1397-1409. doi:doi: 10.4236/psych.2021.129088
- Hulin, C., Drasgow, F., & Parsons, C. (1983). *Item response theory: Application to psychological measurement*. Dallas: Dow Jones-Irwin.
- Hultgard, A. (2009). Odin: An Immigrant to Scandinavia? Dalam H. W. Agneta Ney, *Saga and East Scandinavia: Preprint papers of The 14th International Saga Conference* (hal. 405-410). Gävle: Gävle University Press.
- Huyen, T. (2022, May 27). *5 Types of Exam Questions: Characteristics, Pros & Cons*. Dipetik Oktober 21, 2022, dari Atomi: <https://atomisystems.com/elearning/5-types-of-exam-questions-characteristics-pros-cons/>
- IBM Corp. (2020). IBM SPSS Statistics for Windows (Version 27.0) [Computer software]. IBM Corp.
- Jackson, J. E. (2015). *Oblimin Rotation*. doi: 10.1002/0470011815.b2a13060
- Jimenez, L., & Modaffari, J. (2021, September). *Future of Testing in Education: Effective and Equitable Assessment Systems*. Dipetik Januari 19, 2022, dari Center for American Progress: <https://cfamericanprogress.org/wp-content/uploads/2021/09/EquitableAssesm>
- Johnstone, R. (2001). *Religion and society in interaction: The sociology of religion*. Englewood Cliffs, NJ: Prentice-Hall.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151. doi:<https://doi.org/10.1177/001316446002000116>
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement*, 20, 141-151. doi:[10.1177/001316446002000116](https://doi.org/10.1177/001316446002000116)

- Katzer, J., Cook, K., & Crouch, W. (1982). *Evaluating Information: A Guide for User of Social Science Research*. Reading, MA: Addison-Wesley.
- Kelley, T. (1939). Selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, 30, 17-24.
- Khairuddin, K. N., Omar, Z., Krauss, S. E., & Ismail, I. A. (2018). A Brief Review of the History and Philosophy of Instrument Development in the Social Sciences. *International Journal of Academic Research in Business and Social Sciences*, 8(9), 1517-1524. doi:DOI:10.6007/IJARBSS/v8-i9/4861
- Khan, S. (2014). Qualitative research method: Grounded theory . *International Journal of Business and Management*, 9(11), 224-233. doi:<https://doi.org/10.5539/ijbm.v9n11p224>
- Khan, S., Vijayshri, & Farooqi, F. (2014). A Study of Religiosity in Relation to Spirituality and Anxiety. *International Journal of Advancements in Research & Technology*, 3(4), 269-284.
- Kimberlin, C., & Winterstein, A. (2008). Validity and Reliability of Measurement Instruments Used in Research. *American Journal of Health-System Pharmacists*, 65(1), 2276-2284.
- Koenig, H. G., & Büssing, A. (2010). The Duke University Religion Index (DUREL): A Five-Item Measure for Use in Epidemiological Studies. *Religions* doi:10.3390/rel1010078, 1(1), 78-85. doi:doi:10.3390/rel1010078
- Kucukcan. (2005). Multidimensional Approach to Religion: a way of looking at Religious Phenomenon. *Journal for the Studies of Religion and Ideology*, 4(10), 60-70.
- Kuder, G., & Richardson, M. (1937). The theory of the estimation of test reliability. *Psychometrika*, 151-160. doi:<https://doi.org/10.1007/BF02288391>
- Kuhn, T. S. (1970). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

- Kumar, R. (2019). *Research Methodology, step by step guide for beginner* (ke 5 ed.). London: Sage.
- Lambert, Z. V., Wildt, A. R., & Durand, R. M. (1991). Approximating confidence intervals for factor loadings. *Multivariate Behavioral Research*, 26(3), 421-434. doi:[https://doi.org/10.1207/s15327906mbr2603\\_3](https://doi.org/10.1207/s15327906mbr2603_3)
- Lange, R. (2011). Inter-rater Reliability. Dalam J. Kreutzer, J. DeLuca, & B. Caplan, *Encyclopedia of Clinical Neuropsychology* (hal. 1348). New York: Springer.
- Leach, M. (1990). Philosophical choice. *Journal of Education*, 3(3), 16-18.
- Lenski, G. (1961). *The Religious Factor, A Sociological Study of Religion's Impact on Politics, Economics, and Family Life*. Connecticut: Greenwood Press.
- Likert, R. (1932). Technique for the measurement of attitudes. Dalam *Archives of Psychology* (hal. 1-55).
- Linvall, C. (1969). The review of related research. Dalam W. Gephart, *Educational Research: Selected Reading* (hal. 109-113). Columbus: Charles E. Merril.
- Liu, Y., & Froese., P. (2020). Faith and Agency: The Relationships Between Sense of Control, Socioeconomic Status, and Beliefs About God. *Journal for the Scientific Study of Religion*, 59(2), 311-326. doi:<https://doi.org/10.1111/jssr.12655>
- Livingston, S. (2020). Basic concepts of item response theory: A nonmathematical introduction. Educational Testing Service. Diambil kembali dari [https://www.ets.org/research/policy\\_research\\_reports/publications/report/2020/kbxx.html](https://www.ets.org/research/policy_research_reports/publications/report/2020/kbxx.html)
- Lord, F. (1980). Application of item response theory to practical testing problem. Hillsdale, N.J.
- Louis Cohen, L. M. (2007). *Research Methods in Education*. New York: Routledge.

- Lyman, H. B. (1998). *Test Scores and What They Mean*. Toronto: Allin & Bacon.
- Mack, N., Woodsong, C., Macqueen, K., Guest, G., & Namey, E. (2005). *Qualitative Research Methods: A Data Collector's Field Guide*. Washington: Family Health International. Diambil kembali dari [https://www.researchgate.net/publication/215666086\\_Qualitative\\_Research\\_Methods\\_A\\_Data\\_Collector's\\_Field\\_Guide](https://www.researchgate.net/publication/215666086_Qualitative_Research_Methods_A_Data_Collector's_Field_Guide)
- Magis, D. (2007). Influence, Information and Item Response Theory in Discrete Data Analysis.
- Mari, L., Maul, A., Irribarra, D. T., & Wilson, M. (2017). Quantities, Quantification, and the Necessary and Sufficient Conditions for Measurement. *Meausrement*, 100, 115-121. doi:<https://doi.org/10.1016/j.measurement.2016.12.050>
- Ma'shum, A., & Munawir, Z. A. (1997). *Kamus al-Munawwir, Arab-Indonesia* (ke-3 ed.). Surabaya: Pustaka Progresif.
- Matlock-Hetze, S. (1997). *Basic Concepts in Item and Test Analysis*. Dipetik Juni 22, 2022, dari <https://files.eric.ed.gov/fulltext/ED406441.pdf>
- Matsumoto, D. (2009). *The Cambridge Dictionary of Psychology*. Cambridge: University Press.
- McCall, R. (1970). *Fundamental statistics for psychology*. New York: Harcourt, Brace & World Inc.
- McMillan, J., & Schumacher, S. (1989). *Research in Education, A Conceptual Introduction*. Glenview, IL: Scott, Foresman and Company.
- Meier, B., Fettermann, A., Hauser, D., & Robinson, M. (2020). God is up? Replication and extension attempts of Meier et al. (2007). *Psychology of Religion and Spirituality (Advance online publication)*. doi:<https://doi.org/10.1037/rel0000433>.

- Meijer, R. R., & Tendeiro, J. N. (2018). Unidimensional item response theory. Dalam P. Irwing, T. Booth, & D. J. Hugh, *The Wiley handbook of psychometric testing : A multidisciplinary reference on survey, scale and test development* (hal. 413-433). Wiley. doi:<https://doi.org/10.1002/9781118489772.ch15>
- Messick, S. (1995). Validity of Psychological Assessment Validation of Inferences From Persons' Responses and Performances a Scientific Inquiry Into Score Meaning. *American Psychologist*, 50(9), 741-749.
- Mislevy, R. J. (1995). *Test Theory Reconceived*. Princeton, NJ: Educational Testing Service.
- Mohamed, B., & Sciupac, E. (2018, January 26). *The share of Americans who leave Islam is offset by those who become Muslim*. Diambil kembali dari Pew Research Center: <https://www.pewresearch.org/fact-tank/2018/01/26/the-share-of-americans-who-leave-islam-is-offset-by-those-who-become-muslim/>
- Mueller, D. (1986). *Measuring social attitudes: A handbook for researchers and practitioners*. New York: Teachers' College Press.
- Muijs, D. (2004). *Doing Quantitative Research in Education with SPSS*. London: Sage Publications. Diambil kembali dari [https://www.academia.edu/40027547/Doing\\_Quantitative\\_Research\\_by\\_Daniel\\_Muijs](https://www.academia.edu/40027547/Doing_Quantitative_Research_by_Daniel_Muijs)
- Nafis, M., Hadjar, I., & Hakim, A. (1995). *Keberagamaan Masyarakat Muslim Kodia Semarang*. Semarang: Pusat Penelitian IAIN Walisongo Semarang.
- Nguyen, T. H., Han, H.-R., Kim, M. T., & Chan, K. S. (2014). An Introduction to Item Response Theory for Patient-Reported Outcome Measurement. *Patient*, 7(1), 23–35. doi:[doi:10.1007/s40271-013-0041-0](https://doi.org/10.1007/s40271-013-0041-0)
- Nitko, A. J. (1996). *Educational assessment of students*. Englewood Cliffs, N.J. : Merrill.

- Nudelman, A. E. (1971). Dimensions of Religiosity: A Factor-Analytic View of Protestants, Catholics and Christian Scientists. *Review of Religious Research*, 13(1), 42-56.
- Nunnally, J., & Bernstein, I. (1994). *Psychometric theory*. New York: McGraw-Hill.
- Olufemi, A. (2013). Item Response Theory as a Basis for Measuring Latent Trait of Interest. *Greener Journal of Social Sciences*, 3(7), 378-382. doi:10.15580/GJSS.2013.7.062513691
- Orçan, F. (2018). Exploratory and confirmatory factor analysis: which one to use first? *Journal of Measurement and Evaluation in Education and Psychology*, 9(4), 414-421. doi:10.21031/epod.394323
- Osgood, C. (1952). The nature and measurement of meaning. , 49(3), . *Psychological Bulletin*, 49(3), 197-237. doi: <https://doi.org/10.1037/h0055737>
- Pallant, J. (2011). *A Step by Step Guide to Data Analysis Using the SPSS Program: Survival Manual*. Berkshire : McGraw-Hill.
- Patel, P. (2009, Oktober 15). *Introduction to Quantitative Methods*. Diambil kembali dari Studocu: <https://www.studocu.com/row/document/takoradi-polytechnic/quantitative-studies/quantitative-methods-practical/2647139>
- Pepper, S. C. (1961). *World Hypothesis: A study in Evidence*. London: Cambridge University Press.
- Peraturan Menteri Pendidikan Nasional No. 16. (2006). Peraturan Menteri Pendidikan Nasional Nomor 16 tahun 2007 tentang Standar Kompetensi Guru.
- Perry, S. L., & McElroy, E. E. (2020). Does the Bible Tell Me So? Weighing the Influence of Content versus Bias on Bible Interpretation Using Survey Experiments. *Journal for the Scientific Study of Religion*, 52(2), 311-326. doi:<https://doi.org/10.1111/jssr.12679>
- Popham, W. (1995). *Classroom assessment*. Boston : Allyn and Bacon.

- Popham, W. J. (1981). *Modern educational measurement*. Englewood Clift, NJ: Prentice-Hall.
- Queiros, A., Faria, D., & Almaeda, F. (2017). Strength and limitation of qualitative and quantitative research methods. *European Journal of Education Studies*, 3(9), 369-385.
- Ramlee, N., Osman, A., Salahudinc, S., Yenga, S., Linga, S., & Safizala, M. (2016). The Influence of Religiosity, Stress and Job Attitude towards Organizational Behavior: Evidence from Public Universities in Malaysia. *Procedia Economics and Finance*, 35, 563-573.
- Reise, S., & Haviland, M. (2005). Item Response Theory and the Measurement of Clinical Change. *Journal of Personality Assessment*, 84(3), 228-38. doi:DOI:10.1207/s15327752jpa8403\_02
- Reise, S., & Revicki, D. (2015). Introduction: Age-old problem and modern solution. Dalam S. Reise, & D. Revicki, *Handbook of item response theory modeling: Application to typical performance assessment* (hal. 3-12). New York: Routledge.
- Robson, C. (2011). *Real World Research: A Resource for Users of Social Research Methods in Applied Settings*. Sussex: John Wiley and Sons.
- Rose, D., & Sullivan, O. (1996). *Introducing Data Analysis For Social Scientists*. Buckingham UK: Open University Press.
- Rosenberg, M. (1968). *The logic of survey analysis*. New York: Basic Books Inc.
- Rowntree, D. (1981). *Statistics without Tear: A Primer for Non-mathematicians*. London: Penguin.
- Rowntree, D. (2004). *Statistics without tears : a primer for non-mathematicians*. Boston: Pearson/Allyn & Bacon.
- Salkind, N. J. (2010). *Encyclopedia of Research Design*. New York: Sage.
- Sari, M. (2017). The impact of the age factor on religiosity. *The Journal of International Social Sciences*, 27(2), 257-264.

- Saunders, M. N., Lewis, P., & Thornhill, A. (2016). *Research Methods for Business Students* (7th ed.). Harlow: Pearson Education Limited.
- Schwartz, A. (2007). Social Buffering by God: Can Prayer Reduce Stress, in an Experimental Setting? *Disertasi, Georgia Southern University. Dissertasi, Georgia Southern University.*
- Shandulin, E., & Dmitrieva, P. ( 2021, Mei 20). Testing as a Method of Teaching and Education in the Context of Learning History. *Ural Environmental Science Forum "Sustainable Development of Industrial Region" (UESF-2021), E3S Web of Conferences* 258(07064), 9. doi:<https://doi.org/10.1051/e3sconf/202125807064>
- Sharma, L. R. (2021). Analysis of Difficulty Index, Discrimination Index and Distractor Efficiency of Multiple Choice Questions of Speech Sounds of English. *International Research Journal of MMC (IRJMMC)*, 2(1), 15-28. Diambil kembali dari [www.mmchetauda.edu.np](http://www.mmchetauda.edu.np)
- Shor, E. (2008). Quantification. Dalam W. A. Jr. (Penyunt.), *The International Encyclopedia of the Social Sciences (2nd Edition)* (2 ed., Vol. 6). Farmington Hills, MI: Thomson/Gale.
- Singh, A. S., & Masuku, M. B. (2014). Sampling techniques & determination of sample size in applied statistics research: An overview. *International Journal of Economics, Commerce and Management*, 2(11), 1-22.
- Singh, T., & Kaur, P. (2008). Effect of Meditation on Self Confidence of Student-Teachers in Relation to Gender and Religion. *Journal of Exercise Science and Physiotherapy*, 4(1), 35-43.
- Sodikin, S. A. (2018). Kemampuan guru PAI dalam menyusun instrumen penilaian untuk meningkatkan kemampuan berpikir kritis dan kreatif pada peserta didik di MAN 1 Yogyakarta. *Tesis. UIN Yogyakarta.*
- Spearman, C. (1910). Correlation Calculated with Faulty Data. *British Journal of Psychology*, III, 271-295.

Stevens, S. (1946, June 7). On the Theory of Scales of Measurement. *Science, New Series, Vol. 103, No. 2684 (Jun. 7, 1946), pp. 677-680*, hal. 677-680. Dipetik Maret 4, 2022, dari [https://psychology.okstate.edu/faculty/jgrice/psyc3214/Stevens\\_FourScales\\_1946.pdf](https://psychology.okstate.edu/faculty/jgrice/psyc3214/Stevens_FourScales_1946.pdf)

Sukamolson, S. (2006). Fundamental of Quantitatif Research. Diambil kembali dari <https://www.scribd.com/document/183320041/> suphat-sukamolson

Suryabrata, S. (1991). *Metodologi Penelitian*. Jakarta: Rajawali Pers.

Taglienti, P. (2006). *Dewey Decimal Classification, A Study Manual and Number Building Guide* (22nd ed.). Library Collections, Acquisitions, & Technical Services, 30:3-4, 251,. doi:DOI: 10.1080/14649055.2006.10766133

Tavakol, M., & Wetzel, A. (2020). Factor Analysis: a means for theory and instrument development in support of construct validity. *International Journal of Medical Education*, 6(11), 245-247. doi:10.5116/ijme.5f96.0f4a

Tavakol, M., & Wetzel, A. (2020). Factor Analysis: a means for theory and instrument development in support of construct validity. *International Journal of Medical Education*, 11, 245-247. doi:10.5116/ijme.5f96.0f4a

Tendeiro, J. (2017). *IRT (GMMSGE01): Parametric IRT (dichotomous data)*. Diambil kembali dari [https://bookdown.org/jorgetendeiro/ParametricIRT/\\_main.pdf](https://bookdown.org/jorgetendeiro/ParametricIRT/_main.pdf)

The Mathematic Library. (2021, Juni 6). *LibreText MAthematics*. Diambil kembali dari [https://math.libretexts.org/Bookshelves/Algebra/BEGINNING\\_ALGEBRA/01%3A\\_Real\\_Numbers\\_and\\_The\\_Operations/1.01%3A\\_Real\\_numbers\\_and\\_the\\_Number\\_Line](https://math.libretexts.org/Bookshelves/Algebra/BEGINNING_ALGEBRA/01%3A_Real_Numbers_and_The_Operations/1.01%3A_Real_numbers_and_the_Number_Line)

- The Royal Household at Buckingham Palace. (2021). *Queen, the Church and other faiths*. Diambil kembali dari The Royal Household: <https://www.royal.uk/queens-relationship-churches-england-and-scotland-and-other-faiths>
- Thissen, D., & Orlando, M. (2001). Item response theory for items scored in two categories. Dalam D. Thissen, & W. H., *Test Scoring* (hal. 73-140). Mahwah, NJ:: Lawrence Erlbaum Associates, Inc.
- Thissen, D., & Wainer, H. (2001). *Test scoring*. Mahwah, NJ.: Lawrence Erlbaum.
- Thissen, D., & Wainer, H. (2001). *Test Scoring*. Mahwah, NJ: Lawrence Erlbaum.
- Thompson, N. (2009). *Ability Estimation with Item Response Theory*. Minessota: Assesment System Corporation. USA.
- Tiban, F., Achmad, N., & Zakiyah, S. (2021). Analisis Kualitas Tes Hasil Belajar Matematika Buatan Guru di SMP Negeri 4 Gorontalo. . *Jambura Journal of Mathematics Education*, 2, 70-77.
- Traub, R., & Hambleton, R. (1972). The effect of scoring instructions and degree of speededness on the validity and reliability of multiple-choice tests. *Educational and Psychological Measurement*, 32, 737-758.  
doi:<https://doi.org/10.1177/001316447203200313>
- Trochim, W. M., & Donnelly, J. P. (2008). *The Research Methods Knowledge Base*. Mason, Ohio: Atomic Dog.
- Truncellito, D. (2007). *Epistemology: Internet Encyclopedia of Philosophy*. Diambil kembali dari <http://www.iep.utm.edu/epistemo/>
- Twycross, A., & Shields, L. (2004). Validity and Reliability-What's it All about? Part 2: Reliability in Quantitative Studies. *Paediatric Nursing*, 16(10), 36.
- UCLA: Statistical Consulting Group. (2021). *A practical introduction to factor analysis: Exploratory factor analysis*. Dipetik June 11,

- 2022, dari  
<https://stats.oarc.ucla.edu/spss/seminars/introduction-to-factor-analysis/a-practical-introduction-to-factor-analysis/>
- Undang-undang nomor 12 tahun 2012 tentang Pendidikan Tinggi. (t.thn.).
- van der Linden, W. (2010). Item response theory. Dalam P. Peterson, E. Baker, & McGaw, *International Encyclopedia of Education* (3 ed., hal. 81-89). Monterey, CA.: CTB/McGraw-Hill,. doi:10.1016/B978-0-08-044894-7.00250-5
- Van der Linden, W., & Hambleton, R. (1977). Item response theory: Brief history, common models, and extensions. Dalam W. v. Linden, & R. K. Hambleton, *Handbook of modern item response theory* (hal. 1-28)). New York: Springer.
- van der Linden, W., & Hambleton, R. (1997). Item Response Theory: brief history, common models and extensions . Dalam W. van der Linden, & R. Hambleton, *Handbook of modern item response theory* (hal. 1-28). New York: Springer.
- Verbit, M. (1970). The Components and Dimensions of Religious Behaviour: Toward a Reconceptualization of Religiosity. Dalam P. E. Johnson, *American Mosaic, Social Patterns of Religion in the United States* (hal. 24-37). New York: Random House.
- Verma, J. (2019). *Statistics and Research Methods in Psychology with Excel*. Singapura: Springer. doi:<https://doi.org/10.1007/978-981-13-3429-0>
- Wilandika, A. (2018). Religiosity and Self-Efficacy in the Prevention of HIV-Risk Behaviours among Muslim University Students. *Jurnal Ners*, 13(2), 138-143.
- Williamson, W. P., & Ahmad, A. (2007). Survey Research and Islamic Fundamentalism: A Question about Validity. *Journal of Muslim Mental Health*, 2, 155-176.

Yu, C.-h. (2022). *True score model and item response theory*. Diambil kembali dari [https://www.creative-wisdom.com/teaching/WBI/\\_measurement.shtml](https://www.creative-wisdom.com/teaching/WBI/_measurement.shtml)

Ziegler, M., & Hagemann, D. (2015). Testing the Unidimensionality of Item Pitfalls and Loopholes. *European Journal of Psychological Assessment*, 31(4), 231–237. doi:DOI: 10.1027/1015-5759/a000309

## GLOSARI

**Analisis butir:** Proses mengumpulkan, meringkas, dan menggunakan informasi dari tanggapan testee untuk menilai kualitas pertanyaan dengan jawaban pilihan ganda.

**Analisis faktor R:** Analisis yang digunakan untuk menganalisis hubungan antar variabel untuk mengidentifikasi kelompok variabel pembentuk dimensi (faktor) laten.

**Analisis faktor umum:** Model faktor di mana faktor didasarkan pada matriks korelasi tereduksi. Artinya, kesamaan disisipkan dalam diagonal matriks korelasi, dan faktor yang diekstraksi hanya didasarkan pada varian umum, dengan pengecualian varian spesifik dan galat. Varians umum Varians dibagi dengan variabel lain dalam analisis faktor.

**Analisis kovarian:** analisis staistik yang digunakan untuk menguji signifikansi pengaruh faktor/variabel independen kategoris pada satu variabel dependen dengan kontrol variabel independen kontinum.

**Analisis regresi:** teknik analisis statistik yang digunakan untuk memperediksi nilai suatu matan (variabel dependen/kriteria) berdasarkan nilai amatan lain yang telah diketahui (variabel independen/prediktor) dengan menggunakan persamaan regresi.

**Analisis varian (ANOVA):** teknik analisis statistik yang digunakan untuk menguji perbedaan amatan (satu variabel dependen) antar kelompok (dari satu atau lebih variabel independen/faktor).

**Data:** kumpulan dari serangkaian skor, yang berisi informasi tentang karakteristik suatu variabel dari sekelompok subjek.

**Daya beda butir:** Kemampuan butir untuk membedakan antara kemampuan testee atau peserta tes yang lebih tinggi dan kemampuan mereka yang lebih rendah.

**Daya pengecoh (*distractor*):** Peluang yang menunjukkan apakah masing-masing pilihan jawaban terhadap suatu pertanyaan butir tes mampu berfungsi untuk mengecoh testee yang tidak mengetahui jawabn yang benar.

## INDEX

### **A**

Agama, 9, 11  
Analisis faktor, 121, 146, 147, 150, 151, 152, 156, 159, 160, 166, 272

### **D**

Data, 115  
Daya beda butir, 130, 272

### **I**

Indeks reliabilitas, 87, 273  
Informasi butir, 218

### **K**

Koefisien Alpha, 84, 95  
Koefisien korelasi, 82, 134, 273, 274  
Koefisien reliabilitas, 84, 89, 93, 94, 95, 102, 106, 274  
Komunalitas, 156, 157, 162, 169, 274  
Korelasi, 59, 61, 81, 87, 103, 136, 152, 154, 274, 276, 277

Kurva Karakteristik Butir, 191, 192, 198, 204, 207, 212, 219, 274

### **M**

Masalah, 67, 68, 72, 154  
Metode Belah-Tengah, 92

Metode Kuder-Richardson, 95

Model logistik, 202, 275

### **N**

Nilai Eigen, 163, 275

### **O**

Ogif normal, 194, 275

### **P**

Parameter diskriminasi, 206, 210, 215, 275  
Parameter kesulitan, 200, 275  
Parameter tebakan semu, 211, 214, 276  
Pemuatan faktor, 150, 161, 171, 276  
Pengukuran, 44, 64, 66, 67, 68, 84, 108, 112, 117, 230, 276  
Probabilitas, 183, 192, 193, 205, 237  
Proposal, 129, 183, 191, 193, 198, 202, 206, 210, 232, 239

### **R**

Reliabilitas, 81, 83, 84, 88, 91, 93, 94, 95, 97, 98, 99, 103, 107, 276  
Rerata, 35, 37, 40, 47, 48, 55, 56, 57, 81, 85, 194, 276

Rotasi faktor, 164, 173, 276,  
277

## S

Simpang baku, 51, 53, 277

Skala, 18, 64, 67, 68, 69, 70,  
71, 201, 242, 277, 278

Skala interval, 70, 71, 277

Skala nominal, 69, 277

Skala rasio, 71, 277

Skor baku, 54, 56, 278

Skor galat, 81, 278

Skor komposit, 40, 83, 278

Skor sejati, 13, 78, 79, 80, 81,  
180, 278

Skor tes, 13, 33, 72, 105, 231,  
244, 278

## T

Teori Respons Butir, 14, 180,  
181, 183, 198, 278

Teori tes, 3, 4, 12, 16, 17, 77,  
78, 79, 178, 180, 183, 278

Teori tes klasik, 3, 78, 79, 178,  
180, 278

Teta, 279

Tingkat kesulitan butir, 122,  
200, 279

## U

Uji Kebulatan Bartlett, 155,  
279

Unidimensionalitas, 183, 279

## V

Validitas diskriminan, 122,  
279

Validitas isi, 114, 116, 279

*Validitas konkeren*, 118

Validitas konstruk, 119, 120,  
121, 279, 280

Validitas konvergen, 121, 280

*Validitas kriteria*, 112, 115, 280

*Validitas prediktif*, 117

Validitas tes, 106, 280

Variabel, 58, 66, 80, 146, 148,  
153, 191, 280

## TENTANG PENULIS



Prof. Dr. Ibnu Hadjar, M.Ed., adalah guru besar dalam bidang Ilmu Evaluasi Pendidikan di Fakultas Ilmu Tarbiyah dan Keguruan (FITK) UIN Walisongo Semarang. Menekuni karirnya sebagai dosen sejak lulus dari IKIP Malang (1982). Pendidikan magister ditempuh di Queen's University, Kingston, Canada (1991). Sejak lulus Doktor bidang Penelitian dan Evaluasi Pendidikan di Universitas Negeri Jakarta (2002) juga mengajarkan mata kuliah evaluasi pendidikan pada mahasiswa, baik pada jenjang sarjana, magister, maupun doktor, termasuk di luar alamamaternya, IAIN/UIN Walisongo. Ia juga mengajarkan matakuliah terkait, seperti pengembangan instrumen/tes, metode penelitian kuantitatif, dan lain-lain. Di samping itu, ia juga banyak memberikan pelatihan pada para dosen dan guru.

Di samping mengajar, ia banyak melakukan penelitian, yang hampir seluruhnya memerlukan pengukuran, baik tes maupun kuesener. Karya ilmiahnya banyak ia sebarkan melalui buku, jurnal, dan forum ilmiah, baik di dalam maupun luar negeri. Ia juga membuat Program Aplikasi Statistik Walisongo (W-Stats), termasuk di dalamnya analisis butir tes dan non-tes.

Selama meniti karirnya, ia juga mendapatkan tugas tambahan menjabat sebagai Ketua Jurusan Pendidikan Bahasa Arab (1996-1999), Pembantu Rektor I Bidang Akademik (2003-2007), Dekan Fakultas Tarbiyah (2006-2010), Direktur Program Pascasarjana (2011-2015). Ia juga dipercaya menjadi anggota Majelis Badan Akreditasi Nasional Perguruan Tinggi/BAN-PT (2006-2012) serta anggota Dewan Guru Besar Kementerian Agama.