TEORI PENGUKURAN DALAM PENDIDIKAN

Nonoh Siti Aminah

Program Studi Pendidikan Fisika, PMIPA FKIP, Universitas Sebelas Maret
Surakarta, 57126, Indonesia

Abstrak

Pengukuran dalam pendidikan meliputi pengukuran kemampuan testee dan pengukuran karakteristik alat ukur yang digunakan Ada dua teori pengukuran yang saat ini masih digunakan dan dikembangkan, yaitu teori tes klasik disebut juga classical test theory (CTT) dan teori tes modern disebut juga disebut teori respons item atau item response theory (IRT). Pengukuran menurut teori tes klasik adalahpemberian angka kepada objek atau kejadian dengan aturan tertentu, angka diartikan sebagai sifat yang melekat pada objek. Teori tes klasik menyatakan bahwakarakteristik tesdipengaruhi oleh testee yang menempuh tes tersebut. Jika kelompok peserta yang sama menempuh tes berbeda maka karakteristik kelompok peserta umumnya berubah. Jika diberikan tes yang mudah, kemampuan testee berada pada level tinggi sebaliknya jika diberi kan tes yang sulit, kemampuan testee berada pada level rendah. Menurut teori, kemampuan seseorang tidak berubah karena karakteristik tes. Kelemahan teori tes klasik diatasi oleh teori tes modern atau teori respons item (IRT). Sebagai model alternatif, IRT memiliki sifat antara lain karakteristik item tidak tergantung pada kelompok testee yang dikenai item tersebut atau kemampuan seseorang tidak berubah karena karakteristik item.Model dinyatakan dalam tingkatan item. Model tidak memerlukan tes paralel untuk menghitung koefisien reliabilitas dan model menyediakan ukuran yang tepat untuk setiap kemampuan testee.Parameter statistik pada teori tes klasik yaitu, indeks kesulitan item(b_i) , indeks daya beda item (a_i). Pada teori tes modern, selain indeks kesulitan item dan indeks daya beda item juga ditambah dengan kemampuan peserta tes (θ_j) dan tebakan (c_i).

Kata kunci: Teori pengukuran, teori tes klasik, teori tes modern.

PENDAHULUAN

Pengukuran dalam pendidikan meliputi pengukuran kemampuan testee dan pengukuran karakteristik alat ukur yang digunakan. Karakteristik alat ukur ditunjukkanoleh hasil analisis dari skor hasil pengukuran. Ada dua teori pengukuran yang saat ini masih digunakan dan dikembangkan, yaitu teori tes klasik disebut juga classical test theory (CTT) danteori tes modern disebut juga disebut teori respons item atau item response theory (IRT). Tes yang baik harus memenuhi dua persyaratan yaitu validitas dan reliabilitas.

Validitas mengacu pada seberapa jauh skor suatu tes dapat memberikan bukti bahwa tes telah mengukur konstruk yang didefinisikan. Reliabilitas menunjuk pada kecilnya kesalahan pengukuran. kesalahan pengukuran. Koefisien reliabilitas dinyatakan dengan $\rho_{xx'}$ dankoefisienvaliditasdinyatakan dengan ρ_{XY}.

Bentuk tes ada dua macam yaitu tes objektif dan tes subjektif. Bentuk tes objektif sampai saat ini masih banyak digunakan pada tes formatif dan sumatif. Penggunaan bentuk tes objektif digunakan jika testee relatif banyak serta materi yang diujikan relatif luas, sehingga prinsip objektifitas dapat dipenuhi.

1. Teori Tes Klasik

Teori tes klasik atau teori skor muni didasarkan pada model aditif, yaitu skor amatan atau observed score (X) merupakan penjumlahan dari skor sebenarnya atau true Score (T) dan kesalahan pengukuran atau error (e). Secara matematis, model tersebut Dituliskan $X = T + e$ (Allen dan Yen, 1979: 57). Kesalahan pengukuran bersifat acak. Kesalahan pengukuran merupakan penyimpangan secara teori, antara skor amatan

dengan skor sebenarnya.Pengukuran menurut teori tes klasik adalah pemberian angka kepada objek ataukejadian dengan aturan tertentu (Crocker & Algina, 1986: 3). Angkadiartikansebagaisifat yang melekatpada objek.

Asumsi yang diajukan teori tes klasik ada tujuh yaitu: 1)Skor amatan (X) terdiri dari skor sebenarnya atau true score (T) dan kesalahan pengukuran atau error (e), 2) nilai harapan skor amatan E(X) sama dengan skor sebenarnya (T), E (X) = T. 3) korelasi antara skor sebenarnya dengan kesalahan pengukuran sama dengan nol$\rho_{Te} = 0$. 4) kesalahan pengukuran pada dua tes yang mengukur kemampuan sama tidak berkorelasi$\rho_{e1e2} = 0$. 5) dua tes yang mengukur kemampuan sama, kesalahan pengukuran pada tes pertama (e1) tidak berkorelasi dengan kor sebenarnya pada tes kedua (T2)$\rho_{T2\,e1} = 06$) dua tes yang menghasilkan skor dan memenuhi kelima asumsi pertama disebut tes paralel jika skor sebenarnyadan variansi kesalahan pengukuran yang diperoleh testee setara.7). dua tes yang menghasilkan skor yang memenuhi kelima asumsi pertama disebut essentially τ – equivalent test, jika selisih skor sebenarnya yang diperoleh testee pada tes pertama dan tes kedua merupakan bilangan konstan (Allen & Yen, 1979 : 57). Ketujuh asumsi yang Teori tes klasik menyatakan bahwa karakteristik tes dipengaruhi oleh testee yang menempuh tes tersebut. Jika kelompok peserta yang sama menempuh tes berbeda makakarakteristik kelompok peserta umumnya berubah. (Allen, Yen 1979: 60; DjemariMardapi, 2008:4). Jika diberikan tes yang mudah, kemampuan testee berada pada level tinggi sebaliknya jika diberikan tes yang sulit, kemampuan testee berada pada level rendah. Menurut teori,kemampuan seseorang tidak berubah karena karakteristik tes .

Kelemahan teori tes klasik diatasi oleh teori tes modern atau teori respons item (IRT). Sebagai model alternatif, IRT memiliki sifat antara lain karakteristik item tidak Tergantung pada kelompok testee yang dikenai item tersebut atau kemampuan seseorang tidak berubah karena karakteristik item. Model dinyatakan dalam tingkatan item. Model tidak memerlukan tes paralel untuk menghitung koefisien reliabilitasdan model menyediakan ukuran yang tepat untuk setiap kemampuan testee (Hambleton, Swaminathan & Rogers, 1991: 5).

a. Reliabilitas dan Validitas

Reliabilitas ialah keajegan atau konsistensi hasil pengukuran atau hasil tes yang dilakukan pada waktu yang berbeda dengan subjek sama. Allen dan Yen (1979:72)menyatakan bahwa tes disebut reliabel jika skor amatan mempunyai korelasi yang tinggi dengan skor sebenarnya. Mereka juga menyatakan bahwa reliabilitas merupakan koefisien korelasi antara dua skor amatan yang diperoleh dari hasil pengukuran menggunakan tes paralel. Koefisien reliabilitas (α) suatu tes dapat ditentukan dengan berbagai cara, antara lain, metode belah dua, alfaCronbach, Guttman, dan metode paralel. (Ebel & Frisbie, 1986:231) menyatakan bahwa meskipun tidak ada ketentuan umum, tetapi secara luas dapat diterima bahwa tes yang digunakan untuk membuat keputusan secara perorangan paling tidak harus memiliki koefisien reliabilitas 0,63.

Validitas secara empirik dinyatakan oleh suatu koefisien yang dinamakan koefisien validitas. Koefisien validitas dinyatakan oleh koefisien korelasi antara skor tes yang berkaitan dengan skor kriteria yang relevan. Kriteria berupa skor tes lain yang mempunyai fungsi ukur sama dengan tes yang digunakan atau berupa ukuran lain yang relevan, misalnya tampilan pada suatu pekerjaan, hasil rating yang diperoleh fihak ketiga (Saifuddin Azwar, 2006: 10). Jika skor tes diberi notasi (X) dan skor kriteria dinyatakan dengan (Y), maka koefisien korelasi antara skortes (X) dan skorkriteria (Y)yaituρ_{XY},ρ_{XY} menyatakan tinggi rendahnya validitassuatu alat ukur. Koefisien validitas palingrendah -1 dan paling tinggi 1. Pada kenyataannya sulit untuk mencapai koefisien validitas tertinggi, sebab tidak ada hasil pengukuran yang sempurna.

b. Tingkat Kesulitan

Tingkat kesulitan item dinyatakan dengan proporsi, disimbolkan dengan (p) merupakan salah satu parameter itempada analisis item.Tingkat kesulitan item dapat dihitung dengan berbagai cara, salah satunya proporsi jawaban benar (Bahrul Hayat, 1999). Secara matematis proporsi benar item tertentu(p) dihitung dari banyak testee yang menjawab benar($\sum B$)dibagi banyak testee (N). Secara matematis ditulis, $p = \frac{\sum B}{N}$.

Jika nilai p mendekati 0, maka item terlalu sukar, dan jika nilai p mendekati 1, maka item terlalu mudah. Item yang terlalu mudah atau terlalu sukar tidak memberi informasi yang banyak tentang tes karena tidak mampu membedakan kemampuan testee sehingga perlu dibuang. Allen dan Yen (1979: 121) menyatakan bahwa tingkat kesulitan item (p)yang baik antara 0,3 sampai 0,7 sebagai bentuk informasi tentang kemampuan siswa secara maksimal. Namun angka 0,3 sampai 0,7 perlu disesuaikan dengan tujuan pengembangan item.Setiap pengembangan item mensyaratkan nilai p yang berbeda tergantung pada tujuan pengembangan. Terpenuhinya nilai p yang disyaratkan, akan memperoleh tujuan pengembangan yang maksimal.

c. Daya Beda

Daya beda item merupakan parameter tes yang memberikan informasi tentang seberapa besar item mampu membedakan testee yang kemampuannya tinggi dan testee yang kemampuannya rendah. Daya beda item dihitung menggunakan beberapa cara antara lain koefisien korelasi point biserial (rpbis) dan koefisien korelasi biserial (rbis). Koefisien korelasi point biserialdiperoleh berdasarkan proporsi testee yang menjawab benar pada item (p), mean skor pada tes dari testee yang memiliki jawaban benar pada item ($\bar{Y}_i$) dan mean skor total($\bar{Y}$) dan standard deviation(s_Y) . Ubahan tersebut dinyatakan dalam persamaan $r_{pbis} = \frac{\bar{\bar{Y}}_{i\,-}\ \bar{Y}}{s_Y}\sqrt{\frac{p_X}{1-P_x}}$ dan $r_{bis} = \frac{\bar{\bar{Y}}_{i\,-}\ \bar{Y}}{s_Y}\frac{p_X}{f_Z}$. f (z) adalah ordinat kurva normal dengan z yang diperoleh dari p_X.

Koefisien fungsi korelasi point biserial selalu lebih rendah dibanding dengan koefisien korelasi biserial. Hubungan antara keduanya dinyatakan dengan persamaan, $r_{pbis} = r_{bis}\frac{y}{\sqrt{pq}}$, r_{bis} menyatakan koefisien korelasi biserial, r_{pbis} menyatakan koefisien korelasi point biserial, y menyatakan selisih antara rerata tiap item dengan rerata item total ($\bar{\bar{Y}}_{i\,-}\ \bar{Y}$), p menyatakan proporsi benar itemdan q menyatakan (1 - p).

Kriteria baik tidaknya butir soal menurut Ebel dan Frisbie (1991:234) adalah bila indeks diskriminasi > 0.40 item sangat baik, 0.30 – 0.39) baik tetapi perlu perbaikan, antara 0.20 sampai dengan 0.29 item dengan beberapa catatan, biasanya diperlukan perbaikan, < 0.19 item jelek, dibuang atau diperbaiki melalui revisi.

d. Efektivitas Distraktor

Item pilihan ganda memiliki pengecoh, yaitu jawaban yang tidak bernilai benar. Setiap pengecoh dibuat sedemikian rupa sehingga menarik perhatian testee yang belum memiliki konsep yang baik terhadap materi yang diujikan. Efektifitas distraktor dapat dilihat dari hasil hitung korelasi point biserial darisemua pilihan jawaban yang disediakan. Pada hasil olah menggunakan program ITEMAN pilihan jawaban yang memilki nilai - 9.000 menunjukkan bahwa statistik butir soal atas pilihan jawaban tidak dapat dihitung. Hal ini sering terjadi apabila tidak ada peserta tes yang menjawab butir soal pada pilihan jawaban tersebut (Dikdasmen DikBud: 1999, 116).

e. Kesalahan Pengukuran

Kesalahan standar pengukuran (Standard Error of Measurement: SEM) membantu pemakai tes untuk memahami kesalahan yang bersifat acak. SEM mempengaruhi skor testee. Allen dan Yen (1979: 246), menyatakan bahwa, jika standard deviation dari skor total (s_X),koefisien reliabilitas tes ($\rho_{X\acute{X}}$)maka, kesalahan standar (s_E) dituliskan, $s_E = s_X\sqrt{1 - \rho_{X\acute{X}}}$.

2. Teori Tes Modern

Teori tes modern atau teori respons item (Item Response Theory: IRT), disebut juga teori ciri laten(latent trait theory: LTT), lengkungan karakteristik item atau kurvakarakteristikitem (Item Characteristic Curve), fungsi karakteristik item atau item characteristic function (ICF). Pada makalah ini digunakan satu istilah yaitu teori respons item (Item Responsse Theory: IRT).

Teori respons item (IRT) pada dasarnya memperbaiki kelemahan yang terdapat pada teori tes klasik (CTT), yaitu ketergantungan parameter item kepada kelompok testee dan parameter kemampuan testeekepada parameteritem. Pada IRTparameter item bersifat bebas atau invariance terhadap parameter kemampuan testee. Parameter item dengan parameter kemampuan testeedihubungkan oleh model yang membentuk lengkungan grafik. Parameter item terdiri dari indeks kesulitan item (item difficulty), daya beda item (item discrimination), dan terkaan(guessing). Parameter kemampuan peserta disebut juga ability.

Postulat yang disyaratkan oleh IRT ada dua, yaitu: 1) Kinerja testee pada suatu item dapat diprediksi oleh sekumpulan faktor yang disebut kemampuan. 2) Hubungan antara kinerja testee pada suatu item dan sekumpulan traits dapat digambarkan dalam sebuah fungsi yang disebut fungsi karakteristik item (item characteristic function) atau disebut juga kurva karakteristik item (item characteristic curve), yang disingkat dengan ICC (Hambleton, Swaminathan & Rogers, 1991: 7).

Kurva karakteristik item menyatakan bahwa semakin meningkat level kemampuan seseorang, semakin meningkat pula peluang menjawab benar suatu item tertentu. Kurva karaktristik item menyatakan hubungan sebenarnya antara parameter kemampuan peserta dengan parameter item (Hambleton,Swaminathan & Rogers,1991: 9). Asumsi yang mendasariIRT, yaitu unidimensionalitas danindependensi lokal. Asumsi unidimensionalitas menyatakan bahwa hanya satu kemampuan yang diukur suatu tes.

Persyaratan unidimensi ditujukan untuk mempertahankan invariansi pada IRT. Asumsi lain yaitu, independensi lokal. Independensi lokal dalam praktek sulit dipenuhi, sebab banyak faktor yang mempengaruhi hasil suatu tes. Faktor tersebut antara lain, motivasi, kecemasan, kemampuan untuk bekarja cepat, ketrampilan kognitif di luar kemampuan yang diukur oleh tes. statistik, lokal dimaksudkan sebagai letak pada suatu titik pada parameter kemampuanpeserta (θ). Pada kenyataannya, titik pada parameter kemampuan peserta dapat berbentuk interval. Interval parameter kemampuan peserta diperoleh dari populasi yang homogen atau memiliki kemampuan setara.

Selain homogen, syarat independensi lokal juga menentukan bahwa semua peserta dalam subpopulasi harus independen terhadap item.Dengan kata lain, item memiliki kriteria tidak tergantung pada sampel yang digunakan. Skor pada suatu item tidak bergantung pada skor pada item lain. Independensi lokal disebut juga independensi kondisional (Hambleton, Linden, 1997: 12). Asumsi terpenuhi jika jawaban peserta terhadap suatu item tidak mempengaruhi jawaban terhadap item lain.

Pengujian asumsi dilakukan menggunakan peluang dari pola jawaban setiap testee. Besar peluang sama dengan hasil kali peluang jawaban testee pada setiap item. Hambleton dan Swaminathan (1991) menyatakan bahwa independensi lokal dinyatakan secara matematis dalam persamaan :

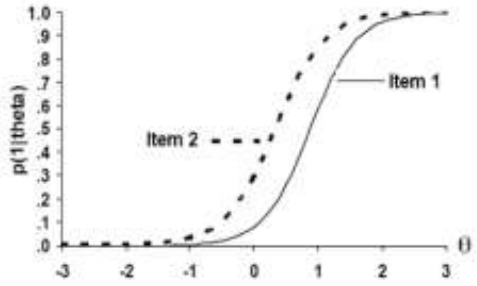
$$p(u_1, u_2, u_3, \dots, \dots, u_n|\theta) = p(u_1|\theta), \dots, \dots, p(u_n|\theta),$$ atau

$$p(u_1, u_2, u_3, \dots, \dots, u_n|\theta) = \prod_{i=1}^{n} P_i(\theta)^{x_i} Q_i(\theta)^{1-x_i}.$$

Persamaan tersebut menghubungkan item (i : 1, 2, ..., n) dengan probabilitas testee yang memiliki kemampuan θ yang dipilih secara acak, dapat menjawab item tertentu dengan benar $p(u_1|\theta)$.

Invariance parameter yaitukemampuan testee tidak berubah karena mengerjakan item yang berbeda tingkat kesulitannya. Demikian juga,parameter item tidak berubah karena diujikan pada kelompok testee yang berbeda tingkat kemampuannya. Hambleton dan Swaminathan (1991) menyatakan bahwa invariance dari parameter kemampuan dapat diselidiki dengan menggunakan dua tes atau lebih yang memiliki tingkat kesulitan berbeda

pada sekelompok testee, seperti ditunjukkan pada Gambar 1.



Gambar 1
Dua Item dengan Indeks Kesulitan Berbeda

IRT memiliki tiga model yaitu, model logistik satu parameter atau juga disebut model Rasch, model logistik dua parameter dan model logistik tiga parameter. Model menunjukkan banyaknya parameter item yang terkandung didalamnya. Model logistik satu parameter memiliki 1 parameter item yaitu indeks kesulitan item (b_i), model logistik duaparameter memiliki dua parameter item yaitu indeks kesulitan item (b_i)dan indeks daya beda item (a_i), model logistik tiga parameter memiliki tiga parameter item, yaitu indeks kesulitan item (b_i), indeks daya beda item (a_i),dan terkaan (c_i).

Secara konseptual, proses pengukuran adalah proses penentuan tempat seseorang pada suatu garis dari variabel yang diukur. Sebelah kiri menunjukkan nilai kurang dan ke kanan menunjukkan nilai lebih. Garis bersifat abstrak, dikonstruksi oleh sejumlahitem dalam tes. Item yang mudah terletak pada bagian sebelah kiri garis, sebaliknya item yang sulit terletak pada sebelah kanan garis.

Syarat yang harus dipenuhi pengukuran yaitu: a) Objek yang diukur satu dimensi, hal ini juga berlaku pada pengukuran sikap. b) Karakteristik suatu objek yang diukur merupakan karakteristik yang dapat dideskripsikan. c) Karakteristik yang akan diukur dapat dideskripsikan oleh pengukuran sebagai besaran linear. d) Satuan ukuran ditentukan oleh suatu proses yang dapat diulang tanpa modifikasi rentang variabel. Esensi dari proses yang dapat diulang tanpa modifikasi adalah model yang menggambarkan bagaimana objek yang diukur dan item berinteraksi untuk menghasilkan observasi yang bermakna.

Hambleton dan Swaminathan (1991)serta Hulin dkk (1983) menyatakan bahwa IRT bertujuan untuk memberikan: 1) Statistik item yang tidak tergantung pada subjek. 2) Skor itemyang menggambarkan kemampuan subjek dan tidak tergantung pada indeks kesulitan item. 3) Model tes dapat memberikan dasar pencocokan antara item dan tingkat kemampuan. 4) Model tes yang asumsi-asumsinya mempunyai dukungan kuat. 5) Model tes tidak memerlukan asumsi paralel dalam pengujian reliabilitas.

a. Item Information Function (IIF)

Item Information Function (IIF) merupakan cara untuk menyatakan kekuatan suatu item pada tes, pemilihan item, dan pembandingan beberapa tes. Fungsi informasi item berkaitan dengan sumbangan item dalam mengungkap latent trait yang diukur dengan tes tersebut. Dengan kata lain, fungsi informasi item memberi informasi tentang kecocokan item dengan model, sehingga membantu seleksi item. Secara matematis fungsi informasi item ke i ialah$I_i(\theta)$ dinyatakan pada Persamaan 2.11. menunjukkan hubungan antara peluang peserta menjawab benar item i dengan kemampuan θ ialah $P_i(\theta)$, turunan $P_i(\theta)$terhadap θ ialah $P_i(\theta)$dan peluang peserta menjawab salah pada item i dengan kemampuan θ ialah $Q_i(\theta)$.$I_i(\theta) = \frac{[P`_i(\theta)]^2}{P_i(\theta)Q_i(\theta)}$ .Menurut Birnbaum (Hambleton dan Swaminathan (1991)), secara matematis fungsi informasi dinyatakan,

$$I_i(\theta) = \frac{2{,}89\, a_i^{\,2}(1 - c_i)}{[(c_i + \exp(Da_i(\theta - b_i)][1 + \exp(-D\, a_i(\theta) - b_i)][1 + ex}$$

Persamaan$I_i(\theta)$ menyatakan hubungan fungsi informasi $I_i(\theta)$ dengan kemampuan subjek(θ), indeks daya beda item ke-i(a_i),indeks kesulitan item ke-i(b_i)dan indeks tebakan item ke-i (c_i)serta bilangan tetap yang besarnya mendekati 2,718.

Berdasarkan persamaan fungsi informasi, fungsi informasi item memiliki sifat: a) Pada respons item model logistik fungsi informasi item mendekati maksimal ketika (b_i) mendekatiθ.Pada model logistik tiga parameter fungsi informasi maksimal dicapai ketika θ terletak sedikit di atas (b_i)dan tebakan item menurun. b) Fungsi informasi secara keseluruhan meningkat jika (b_i)meningkat. Fungsi informasi tes merupakan jumlah dari fungsi informasi item penyusun tes tersebut (Hambleton dan Swaminathan, 1991), sehingga fungsi informasi tes akan besar jika item penyusun tes mempunyai fungsi informasi yang besar. Fungsi informasi dinyatakan dalam persamaan$I(\hat{\theta}) = \sum_{i=1}^{n} I_i(\theta)$.$I(\hat{\theta})$ menyatakan estimasi fungsi informasi tes, $\sum_{i=1}^{n} I_i(\theta)$ menyatakan jumlah fungsi informasi tiap item.

Kesalahan standar pengukuran atau standard error measurement (SEM), berkaitan dengan fungsi informasi, yaitu berbanding terbalik secara kuadratik. Makin besar fungsi informasin SEM makin kecil atau sebaliknya (Croker dan Algina, 1996). Jika fungsi informasi dinyatakan dengan I(θ) dan SEM dinyatakan dengan SEM (θ) maka secara matematis hubungan keduanya dinyatakan pada persamaan SEM $(\theta) = \frac{1}{\sqrt{I(\hat{\theta})}}$.

Efisiensi relatif (ER) adalah perbandingan nilai fungsi informasi dari dua tes yang berbeda. Jika fungsi informasi tes A dinyatakan dengan $I_A(\hat{\theta})$ dan fungsi informasi tes B dinyatakan dengan$I_B(\hat{\theta})$ maka efisiensi relatif atau dinyatakan dengan persamaan ER $(\theta) = \frac{I_A(\hat{\theta})}{I_B(\hat{\theta})}$. Hasil kajian tentang CTT dan IRT ditinjau dari item statistik diringkas seperti pada Tabel 1.

Tabel 1
Statistik Item pada Teori Tes Klasik dan Teori Respons Item

Statistik	Teori Tes Klasik	Teori Tes Modern
Indeks Kesulitan item (b_i)	proporsi benar dari testee (p) tergantung pada sampel.	Parameter b_i, lokasi item pada skalaindekskesulitan, dapat dideteksi menggunakan model logistiksatu parameter.
Indeks Daya Beda item (a_i)	*biserial* dari atau *point biserial*, korelasi antara skor item dan skor total	Parameter a_i , merupakan slope dari garis singgung pada titik b itemtertentu, dapat dideteksi menggunakan model logistik dua parameter.
Kemampuan Peserta tes (θ_j)	Skor amatan dari subjek yangmengikuti tes.	Estimasi parameter θ.
Akurasi Skor	Reliabilitas dan *standard errorof measurement* (SEM) darirerata skor tes.	*Standard error* dari estimasi kemampuan.
Tebakan (c_i)	Tidak terdeteksi	Parameter c_i (*guessing*), hanya dapat dideteksi jika menggunakan model logistik tiga parameter.

DAFTAR PUSTAKA

Allen, M.J. & Yen,W.M. (1979). Introduction to measurement theory. Monterey, CA: Brooks/ Cole Publishing Company.

Crocker & Algina. (1986), Introduction to classical and modern test theory. New York: United State of America: CBS College Publishing.

Djemari Mardapi. (2004). Penyusunan tes hasil belajar. Yogyakarta: Program Pascasarjana Universitas Negeri Yogyakarta.

Djemari Mardapi.(2008). Teknik penyusunan instrumen tes dan non tes. Yogyakarta: Mitra Cendikia Press.

Dikdasmen Dikbud (1999). Pengelolaan pengujian bagi guru mata pelajaran. Jakarta: Departemen Pendidikan dan Kebudayaan Direktorat Jenderal Pendidikan Dasar dan Menengah Direktorat Pendidikan Menengah Umum.

Embretson & Reise. (2000), Item response theory for psychologists. London : Lawrence Erlbaum associates publishers.

Ebel R.L. & Friesbie. D.A (1986), Essentials of educational measurement. New Jersey: Prentice – Hall. Inc.

Hambleton, R. K., Swaminathan, H., & Rogers, H.J. (1991). Fundamental of item response theory. Newbury Park, CA: Sage Publication. New Inc.

Hambleton, R.K. & Swaminathan, H. (1985). Item response theory principles and applications. Boston, MA: Kluwer Inc.

Nonoh Siti Aminah. (2011). Karakteristik metode penyetaraan skor tes untuk data dikotomos. Disertasi tidak diterbitkan.