



**SEMINAR NASIONAL KIMIA DAN PENDIDIKAN KIMIA VIII**  
“Peningkatan Profesionalisme Pendidik dan Periset Sains Kimia di  
Era Masyarakat Ekonomi Asean (MEA)”  
Program Studi Pendidikan FKIP UNS  
Surakarta, 14 Mei 2016



**MAKALAH  
PENDAMPING**

**PARALEL B**

**ISBN : 978-602-73159-1-4**

**PENINGKATAN KOMPETENSI CALON PENDIDIK KIMIA MELALUI  
ITEM RESPONSE THEORY: STRATEGI  
MENGHADAPI MASYARAKAT EKONOMI ASEAN**

**Kriswantoro<sup>1\*</sup>, Rizki Nor Amelia<sup>1</sup>, Irwanto<sup>2,3</sup>**

<sup>1</sup>Magister Penelitian dan Evaluasi Pendidikan, Program Pascasarjana, Universitas Negeri  
Yogyakarta, Yogyakarta, Indonesia

<sup>2</sup>Magister Pendidikan Kimia, Program Pascasarjana, Universitas Negeri Yogyakarta, Yogyakarta,  
Indonesia

<sup>3</sup>SMA Tiga Maret Yogyakarta, Indonesia

\*Keperluan korespondensi, telp: 081373298561, kriswantoro18@gmail.com

**ABSTRAK**

Artikel ini bertujuan untuk memberikan gambaran tentang urgensi *Item Respons Theory* (IRT) sebagai alat analisis yang terstandarisasi secara internasional. Pentingnya penggunaan IRT perlu ditekankan, mengingat analisis menggunakan pendekatan ini akan menghasilkan parameter butir yang terstandar, sehingga kemampuan subjek (siswa) yang diukur dapat diperbandingkan dan diinterpretasi dengan mudah. Desain penelitian ini merupakan studi literatur, dengan menganalisis kemungkinan kekurangan yang dimiliki oleh calon-calon pendidik kimia di Indonesia. Hal ini diperlukan karena dalam menghadapi persaingan dunia kerja pada era Masyarakat Ekonomi Asean (MEA), pendidikan kimia dituntut dapat meluluskan calon-calon pendidik yang dapat bersaing secara kompetitif. Penelitian ini menyimpulkan bahwa salah satu bekal dalam bersaing secara internasional, calon pendidik kimia wajib memiliki kompetensi untuk membuat instrumen pengukuran kimia yang terstandar. Instrumen terstandar dapat memberikan kesalahan pengukuran yang minimal, sehingga hasil pengukuran menjadi lebih cermat.

**Kata kunci:** *calon pendidik kimia, Item Respons Theory, Masyarakat Ekonomi Asean*

**PENDAHULUAN**

Terhitung sejak tahun 2015, *Asean Economic Community* (AEC) atau dalam istilah bahasa Indonesia adalah Masyarakat Ekonomi Asean (MEA) resmi diberlakukan. Meski belum begitu terasa dampaknya, namun lambat laun MEA akan merubah tatanan ekonomi secara global. MEA merupakan suatu bentuk kerja sama antara anggota negara-negara ASEAN

yang terdiri dari Brunei, Filipina, Indonesia, Kamboja, Laos, Myanmar, Singapura, Thailand, dan Vietnam. Melalui kerja sama tersebut akan diberlakukan perdagangan bebas antara negara-negara ASEAN dengan cara membentuk pasar tunggal yang terdiri dari 600 juta konsumen. Intinya, MEA dipersiapkan dan dirancang untuk mewujudkan wawasan ASEAN 2020. Dengan diberlakukannya MEA,

jelas persaingan usaha akan semakin sempit, walau pangsa pasar menjadi lebih luas karena meliputi negara-negara ASEAN.

Dalam menghadapi persaingan yang teramat ketat selama MEA ini, negara-negara ASEAN, tidak terkecuali Indonesia, haruslah mempersiapkan sumber daya manusia (SDM) yang terampil, cerdas, dan kompetitif. Semua produktivitas lintas sektor harus ditingkatkan, terutama sektor pendidikan. Sektor ini perlu diperhatikan mengingat baik buruknya kualitas tenaga kerja yang dihasilkan merefleksikan kualitas pendidikan yang telah dikenyam. Kualitas pendidikan yang tidak segera diperbaiki, akan menjadi ancaman bagi pelajar dan lulusan Indonesia. Mereka tidak dapat bersaing dengan lulusan negara lain yang kualitas pendidikan sudah jauh lebih mapan, misalnya Negara Vietnam. Wardiman menuturkan bahwa upah minimum regional di Indonesia tercatat lebih besar daripada Vietnam, sehingga kemungkinan negara Indonesia didatangi para pekerja Vietnam sangatlah besar [1]. Dari pendapat tersebut dapat disimpulkan bahwa pemerintah perlu segera memperkuat tenaga pendidik (pengajar) Indonesia, karena merekalah yang kemudian akan mempersiapkan pelajar dalam negeri agar unggul bersaing di lingkup ASEAN.

Untuk menciptakan peserta didik yang berkualitas, baik calon pendidik maupun guru harus menguasai empat kompetensi, yaitu kompetensi pedagogik, profesional, sosial, dan kepribadian [2]. Pertama, kompetensi pedagogik. Kompetensi menuntut kemampuan guru untuk mampu mengelola pembelajaran. Kompetensi ini merupakan kompetensi khas karena membedakan guru dengan profesi lainnya. Kompetensi pedagogik terdiri dari 7 aspek kemampuan, yaitu: (a) mengenal

karakteristik anak didik; (b) menguasai teori belajar dan prinsip-prinsip pembelajaran; (c) mampu mengembangkan kurikulum; (d) kegiatan pembelajaran yang mendidik; (e) memahami dan mengembangkan potensi peserta didik; (f) komunikasi dengan peserta didik; dan (g) penilaian dan evaluasi pembelajaran.

Kedua, kompetensi profesional. Kompetensi ini dapat dilihat dari kemampuan guru dalam mengikuti perkembangan ilmu terkini mengingat perkembangan ilmu bersifat dinamis. Kompetensi profesional menuntut guru untuk menguasai materi pembelajaran secara luas dan mendalam yang meliputi: (a) konsep, struktur, metode keilmuan atau teknologi atau seni yang menaungi (koheren) dengan materi ajar; (b) materi ajar yang ada dalam kurikulum sekolah; (c) hubungan konsep antar pelajaran terkait; (d) penerapan konsep-konsep keilmuan dalam kehidupan sehari-hari; dan (e) kompetensi secara profesional dalam konteks global dengan tetap melestarikan nilai dan budaya nasional.

Ketiga, kompetensi sosial. Kompetensi sosial bisa dilihat dari interaksi guru dalam bermasyarakat dan bekerja sama dengan peserta didik serta guru-guru lainnya. Kompetensi sosial yang harus dikuasai guru meliputi: (a) berkomunikasi lisan dan tulisan; (b) menggunakan teknologi komunikasi dan informasi secara fungsional; (c) bergaul secara efektif dengan peserta didik, sesama pendidik, tenaga kependidikan, orang tua/wali peserta didik; (d) bergaul secara santun dengan masyarakat sekitar; (e) bertindak sesuai norma agama, hukum, sosial, dan kebudayaan nasional Indonesia; (f) menunjukkan pribadi yang dewasa dan teladan; (g) etos kerja, tanggung jawab yang tinggi, rasa bangga

menjadi guru. Terakhir, kompetensi kepribadian. Kompetensi ini terkait dengan guru sebagai teladan, beberapa aspek kompetensi ini misalnya: (a) dewasa; (b) stabil; (c) arif dan bijaksana; (d) berwibawa; (e) mantap; (f) berakhlak mulia; (g) menjadi teladan bagi peserta didik dan masyarakat; (h) mengevaluasi kinerja sendiri; (i) mengembangkan diri secara berkelanjutan.

Berdasarkan paparan kompetensi yang telah dijelaskan di atas, maka artikel hanya ini akan berfokus pada kompetensi pedagogi yang wajib dimiliki oleh guru maupun calon pendidik, terutama yang berkaitan dengan penilaian dan evaluasi pembelajaran. Penilaian didefinisikan sebagai suatu proses untuk mendapatkan informasi yang digunakan untuk membuat keputusan tentang siswa, kurikulum, program, sekolah, dan kebijakan pendidikan [3]. Keputusan-keputusan yang berkaitan dengan siswa dapat meliputi penempatan siswa pada program pendidikan yang berbeda, pemberian nilai pada siswa, pembimbingan siswa, pemilihan siswa untuk mengikuti program-program pendidikan, pemberian penghargaan dan sertifikat terhadap kompetensi siswa. Keputusan tentang kurikulum, program, dan sekolah menyangkut bagaimana efektivitas dan cara memperbaiki hal-hal tersebut. Selanjutnya, keputusan tentang kebijakan pendidikan dikaitkan dengan pengambilan keputusan di tingkat sekolah, tingkat daerah, dan tingkat pusat. Sementara evaluasi pembelajaran sebenarnya lebih merujuk pada evaluasi hasil belajar. Evaluasi hasil belajar didefinisikan sebagai suatu rangkaian kegiatan untuk mendapatkan informasi, dimana informasi tersebut digunakan sebagai bahan pertimbangan dalam membuat keputusan tentang pencapaian hasil belajar

siswa terhadap materi yang sudah diajarkan [4].

Evaluasi hasil belajar dapat dikategorikan dalam dua kelompok yaitu formatif dan sumatif [5]. Evaluasi formatif bertujuan untuk memperoleh masukan tentang tingkat keberhasilan pelaksanaan proses pembelajaran. Masukan ini berguna untuk memperbaiki strategi pembelajaran. Tes formatif dilakukan secara periodik sepanjang semester. Materi tes dipilih berdasarkan standar kompetensi dan kompetensi dasar. Sementara itu, evaluasi sumatif diberikan di akhir suatu pelajaran atau semester. Hasilnya digunakan untuk menentukan keberhasilan belajar siswa pada pelajaran tertentu. Tingkat keberhasilan ini dinyatakan dengan skor atau nilai, pemberian sertifikat, dan sejenisnya. Tingkat kesukaran soal dalam tes sumatif bervariasi, sedangkan materinya harus mewakili bahan yang diajarkan [6].

Membahas tentang penilaian dan evaluasi, tidak lengkap jika tidak diiringi dengan pembahasan pengukuran. Hal ini dikarenakan pengukuran merupakan pijakan utama yang harus dilakukan sebelum melakukan penilaian maupun evaluasi. Pengukuran dalam bidang pendidikan, terutama bidang pendidikan kimia, adalah salah satu cara pengumpulan informasi dalam bidang pembelajaran kimia yang hasilnya dapat dikuantifikasikan atau dinyatakan dalam bentuk angka [7], yang selanjutnya disebut skor. Teknik pengukuran hasil belajar kimia berupa ujian atau non ujian. Instrumen pengukuran hasil belajar kimia berupa soal, non soal, atau tugas-tugas. Instrumen yang digunakan haruslah sah (*valid*) dan konsisten (*reliable*) agar diperoleh hasil pengukuran yang baik. Namun, selain syarat *valid* dan *reliable*, parameter butir yang terkandung dalam

instrumen pengukuran kimia juga harus terstandar. Salah satu cara melakukan standardisasi parameter butir, diperlukan Item Respons Theory (IRT). Parameter butir yang terstandar memiliki makna bahwa butir memiliki karakteristik psikometris yang baik, sehingga jika dilakukan pengukuran menggunakan butir ini, maka hasil pengukuran akan lebih cermat. Hal ini perlu ditekankan mengingat berbagai keputusan penting yang diperoleh, didasarkan pada skor peserta tes dari hasil mengerjakan instrumen tersebut [8]. Aplikasi lebih jauh terkait IRT akan sangat bermanfaat, terutama pada analisis karakteristik butir soal, pengembangan instrumen (baik kognitif, afektif, maupun psikomotorik), penyetaraan skor tes, hingga identifikasi bias butir.

## METODE PENELITIAN

Desain penelitian ini merupakan studi literatur, dengan menganalisis kemungkinan kekurangan yang dimiliki oleh calon-calon pendidik kimia di Indonesia. Hal ini diperlukan karena dalam menghadapi persaingan dunia kerja pada era Masyarakat Ekonomi Asean (MEA), pendidikan kimia dituntut dapat meluluskan calon-calon pendidik yang dapat bersaing secara kompetitif.

## HASIL DAN PEMBAHASAN

### 1. Latar Belakang Munculnya IRT

Dalam pengukuran pendidikan, terdapat dua pendekatan yang sering digunakan yaitu *Classical Test Theory* (CTT) dan *Item Respons Theory* (IRT). Bila menggunakan CTT, umumnya siswa menjawab butir soal suatu tes yang berbentuk pilihan ganda dengan benar diberi skor 1 dan 0 jika salah, sehingga kemampuan siswa dinyatakan dengan skor total yang dipe-rolehnya. Prosedur tersebut

kurang memperhatikan interaksi antara setiap orang siswa dengan butir. Namun, pendekatan IRT merupakan pendekatan alternatif yang dapat digunakan dalam menganalisis suatu tes. Hal ini dikarenakan IRT menggunakan model probabilistik.

Model probabilistik bermakna bahwa probabilitas subjek untuk menjawab butir dengan benar bergantung pada kemampuan subjek dan karakteristik butir. Artinya, peserta tes berkemampuan tinggi mempunyai probabilitas menjawab benar lebih besar dibandingkan peserta tes yang berkemampuan rendah. CTT memiliki beberapa kelemahan yaitu: (1) tingkat kesukaran dan daya beda butir soal tergantung pada kelompok peserta yang mengerjakannya, (2) penggunaan metode dan teknik untuk desain dan analisis tes dengan memperbandingkan kemampuan siswa pada pembagian kelompok atas, tengah, dan bawah, (3) Konsep reliabilitas skor didefinisikan dari istilah tes paralel, (4) tidak ada dasar teori untuk menentukan bagaimana peserta memperoleh tes yang sesuai dengan kemampuan peserta yang bersangkutan, dan (5) *Standard Error Measurement* (SEM) berlaku pada seluruh peserta tes [9]. Berdasarkan kelemahan-kelemahan tersebut, maka pendekatan IRT muncul untuk mengatasi kelemahan yang ada pada CTT.

### 2. Tujuan IRT

IRT dikenal juga sebagai Teori Ciri Laten (*Latent Trait Theory* -LTT) atau lengkungan karakteristik butir (*Item Characteristic Curve*-ICC) atau Fungsi Karakteristik Butir (*Items Characteristic Function*-ICF) [10]. Pada dasarnya, teori ini ingin memperbaiki kelemahan yang terdapat pada CTT yakni adanya sifat *group dependent* dan *item dependent*. Hal ini berarti indeks daya

pembeda, tingkat kesulitan, dan koefisien reliabilitas tes tergantung kepada yang mengerjakan tes tersebut, selain dipengaruhi oleh soal atau butir yang ada [11].

Untuk mencapai tujuan seperti ini, IRT membangun suatu model yang menghubungkan ciri butir dengan ciri peserta. Dengan sejumlah syarat tertentu, model hubungan itu dibuat untuk berlaku secara bebas bagi kelompok butir dan kelompok peserta mana saja yang memenuhi syarat itu. Dengan kata lain, model hubungan tersebut dibuat untuk berlaku bagi sejumlah kelompok butir dan sejumlah kelompok peserta tanpa ketergantungan satu terhadap ciri lainnya. Ciri butir dan ciri peserta yang dihubungkan oleh model yang berbentuk fungsi atau lengkungan grafik dengan sejumlah syarat itu dinyatakan melalui sejumlah parameter. Ada parameter ciri butir dan ada pula parameter ciri peserta dengan menggunakan cukup banyak butir tes serta cukup banyak respon peserta tes, dari model hubungan itu kita dapat mengestimasi parameter ciri butir dan parameter ciri peserta. Demikianlah dengan adanya butir tes, peserta tes, respon peserta, ciri butir, ciri terpendam (laten) dari peserta, model hubungan berbentuk fungsi atau berbentuk lengkungan grafik ini, muncullah berbagai istilah seperti teori respon butir, teori ciri laten, lengkungan karakteristik butir, dan fungsi karakteristik butir. Sesuai dengan namanya, teori respon butir membangun model hubungan untuk setiap butir yakni hubungan diantara butir itu dengan para peserta yang meresponsnya.

Terdapat tiga unsur utama dalam IRT. Unsur pertama adalah butir. IRT menelaah butir untuk menemukan cirinya. Setiap penelaahan berkenaan dengan satu butir sehingga di dalam tes, kita dapat memiliki banyak penelaahan,

masing-masing berkenaan dengan satu butir di dalam perangkat tes itu. Unsur kedua adalah peserta yang merespons butir itu. Biasanya, peserta yang merespons butir itu melakukan responsinya melalui suatu kemampuan. Dalam hal ini, model pada IRT menggunakan suatu skala kontinum untuk menampung segala jenis kemampuan peserta yang merespons butir. Diangkat ke istilah yang lebih umum, kontinum segala kemampuan peserta, dinamakan kontinum ciri peserta atau dalam sejumlah hal, dinamakan juga kontinum ciri terpendam (*latent trait*) peserta. Unsur ketiga adalah isi responsi peserta terhadap butir tes. Didalam tes, isi respon dapat berbentuk salah atau benar, sehingga dengan menggabungkan isi reponsi dari semua peserta terhadap butir itu, kita menemukan hasil berupa banyaknya jawaban benar, proporsi jawaban benar, atau probabilitas jawaban benar. Dengan cara yang sama, kita juga dapat menemukan banyaknya jawaban salah, proporsi jawaban salah, atau probabilitas jawaban salah.

### 3. Persyaratan dan Hakikat IRT

Ada tiga pokok persyaratan (asumsi) dalam IRT yaitu unidimensi, independensi lokal, dan invariansi parameter [10]. Pada umumnya, IRT mensyaratkan bahwa setiap butir hanya mengukur satu ciri di kalangan peserta, meskipun belakangan ini berkembang IRT Multidimensi. Namun pembahasan kita disini, IRT dibatasi pada butir unidimensi sehingga unidimensi inilah yang menjadi salah satu syarat pada butir itu. Persyaratan butir unidimensi ditujukan untuk mempertahankan invariansi pada IRT. Jika suatu butir tes mengukur lebih dari satu dimensi, maka jawaban terhadap butir itu merupakan kombinasi dari berbagai kemampuan peserta. Akibatnya kita tidak lagi mengetahui kontribusi

dari setiap kemampuan terhadap jawaban peserta. Misalnya suatu tes bertujuan untuk mengetahui kemampuan kimia pada materi tertentu. Jika kita tidak hati-hati dalam mengkonstruksi butir, selain mengukur kimia, butir juga bisa mengukur kemampuan berbahasa karena butir diungkapkan melalui bahasa. Sekiranya siswa memberi jawaban salah, maka kita tidak lagi mengetahui apakah kesalahan disebabkan oleh kemampuan peserta di bidang fisika atau di bidang bahasa.

Selain unidimensi, IRT juga mensyaratkan adanya independensi lokal. Disini, lokal dimaksudkan sebagai letak pada suatu titik di kontinum ciri peserta  $\theta$ . Pada prakteknya, titik pada kontinum peserta dapat berbentuk interval. Dan di dalam titik atau di dalam interval parameter ciri peserta itu terhadap subpopulasi yang homogen. Di

samping homogen, syarat independensi lokal menentukan bahwa semua peserta di dalam subpopulasi itu harus independen terhadap butir tes. Ini berarti bahwa dengan independensi lokal, skor dari sejumlah butir tes yang dijawab oleh subpopulasi yang sama, masing-masing haruslah independen. Dengan demikian, skor dari satu butir tes tidak boleh ditentukan atau bergantung kepada skor pada butir tes yang lain. Syarat terakhir adalah invariansi parameter. Melalui syarat pokok ini, kita menemukan bahwa fungsi atau lengkungan responsi atau karakteristik butir adalah tetap atau tidak berubah sekalipun kelompok peserta yang menjawab butir yang sama itu berubah-ubah. Dan untuk kelompok yang sama, ciri mereka adalah tetap sekalipun butir yang mereka jawab berubah-ubah.

#### 4. Model-Model IRT

**Tabel 1. Jenis data dan Model yang digunakan dalam IRT**

No.	Jenis Data	Model yang digunakan	Referensi
1.	Dichotomous	Latent Linear	Lazarsfeld & Henry (1968)
		Perfect Scale	Guttman (1944)
		Latent Distance	Lazarsfeld & Henry (1968)
		One-, Two-, Three-Parameter Normal Ogive	Lord (1952)
2.	Multicategory Scoring	One-, Two-, Three-Parameter Logistic	Birbaum (1957, 1958 <sup>a</sup> , 1958 <sup>b</sup> , 1968), Lord & Novick (1968). Lord (1980 <sup>a</sup> ), Rasch (1960), Wright & Stone (1979)
		Four-Parameter Logistic	McDonald (1967), Barton & Lord (1981)
		Nominal Response	Bock (1972)
3.	Continuous	Graded Response	Samejima (1969)
		Partial Credit Model	Master (1982)
3.	Continuous	Continuous Response	Samejima (1972)

Tabel 1 menunjukkan bahwa terdapat tiga jenis sistem penskoran yang terdiri dari sistem penskoran *dichotomous*, *multicategory*, dan *continuous* berdasarkan jenis data yang dimiliki. Dari ketiga sistem tersebut, penskoran *dichotomous* adalah yang paling umum digunakan dalam bidang penilaian pendidikan. Model ini sendiri memiliki bermacam-macam tipe seperti *true-false*, *short answer*, *sentence completion*, dan *matching* [9]. *Dichotomous IRT models* hanya cocok untuk item yang memiliki dua kategori skor yang mungkin seperti kategori benar-salah. Sedangkan, untuk item yang memiliki lebih dari dua kategori skor, model yang sesuai ialah sistem penskoran *multicategory*. Sebagai

contoh untuk model ini ialah item tes yang memberikan kredit parsial, seperti pertanyaan essay yang skornya diberi rating mulai dari nol hingga empat (0 – 4). Selain itu, model ini juga dapat berupa item survei dengan tingkat respons yang beragam seperti sangat tidak setuju, tidak setuju, setuju, atau sangat setuju. Sedangkan, sistem penskoran yang terakhir yaitu *continuous scoring system* merupakan model yang jarang digunakan serta kurang dikenal oleh para praktisi. Salah satu penyebab ketidakpopuleran *continuous scoring system* ialah kekurangan software yang mudah diakses untuk mengestimasi model parameternya

**Tabel 2. Rumus Matematis Model Logistik Data Dikotomus**

Model	Rumus matematis
1-PL	$P_i(\theta) = \frac{e^{D(\theta-b_i)}}{1 + e^{D(\theta-b_i)}} \quad [1 + e^{-D(\theta-b_i)}]^{-1}$
2-PL	$P_i(\theta) = \frac{e^{D_{ai}(\theta-b_i)}}{1 + e^{D_{ai}(\theta-b_i)}} \quad [1 + e^{-D_{ai}(\theta-b_i)}]^{-1}$
3-PL	$P_i(\theta) = c_i + (1 - c_i) \frac{e^{D_{ai}(\theta-b_i)}}{1 + e^{D_{ai}(\theta-b_i)}} \quad c_i + (1 - c_i)[1 + e^{-D_{ai}(\theta-b_i)}]^{-1}$
4-PL	$P_i(\theta) = c_i + (\gamma_i - c_i) \frac{e^{D_{ai}(\theta-b_i)}}{1 + e^{D_{ai}(\theta-b_i)}} \quad c_i + (\gamma_i - c_i)[1 + e^{-D_{ai}(\theta-b_i)}]^{-1}$

Pada prinsipnya, IRT menggunakan distribusi normal. Namun penghitungan menggunakan distribusi normal agak rumit, sehingga digunakanlah distribusi logistik [6]. Dengan menggunakan distribusi logistik, maka model pada IRT berdasarkan jumlah parameter butirnya dibagi menjadi 4, yaitu *one-parameter logistic model* (1-PL), *two-parameter logistic model* (2-PL), *three-parameter logistic model* (3-PL), dan *four-parameter logistic model* (4-PL) [9]. Namun, dari keempat model tersebut model 1-PL, 2-PL, dan 3-PL merupakan model

yang umumnya digunakan. Ketiganya dibedakan melalui jumlah parameter yang dimiliki masing-masing model sebagai pembentuk fungsi respons item. Saat jumlah parameter dalam tiap model bertambah (contohnya, dari 1 ke 2 ke 3), model tersebut menjadi lebih *flexible* dan memberikan refleksi yang lebih realistis mengenai bagaimana respons yang diharapkan kepada tiap item dihubungkan dengan kemampuan yang mendasarinya. Bagaimanapun juga, dalam konteks praktis tiap-tiap model memiliki

keuntungan masing-masing sehingga ketiganya digunakan secara luas dalam aplikasi pengujian. Berikut sedikit ulasan mengenai model 1-4 PL, sedangkan model matematisnya dirangkum pada Tabel 2.

**a. One-parameter logistic model (1-PL)**

Model 1-PL adalah model IRT yang paling sederhana dimana hanya terdapat satu item parameter. Parameter yang dimaksud ialah tingkat kesukaran item yang dilambangkan dengan huruf  $b$ . Pada saat nilai  $b$  meningkat, ini berarti tingkat kesukaran item juga meningkat. Pada saat skor kemampuan dari satu kelompok peserta ujian ditransformasikan maka rata-ratanya ialah nol (0) dan standar deviasinya ialah satu (1). Indeks tingkat kesukaran biasanya berkisar antara kira-kira -2,0 logit hingga +2,0 logit [9]. Nilai yang semakin mendekati -2 logit menunjukkan karakteristik butir yang semakin mudah, dan nilai yang mendekati +2 logit menunjukkan karakteristik butir yang semakin sulit bagi peserta ujian. Dalam model 1-PL, nilai  $b$  melambangkan tingkat kemampuan ( $\theta$ ) yang dituntut bagi 50% kesempatan menjawab dengan tepat. Jadi, jika  $b = 0$  logit, maka kemungkinan jawaban benar akan sama dengan 0,5 pada tingkat kemampuan  $\theta = 0$  logit [12].

**b. Two-parameter logistic model (2-PL)**

Model 2-PL memiliki keuntungan dalam hal kemudahan, namun model ini kekurangan fleksibilitas yang membolehkan butir yang berbeda memiliki fungsi respons butir dari kemiringan atau kecuraman yang berbeda. Model 2-PL mengatasi keterbatasan model ini dengan menambahkan parameter kedua (dilambangkan dengan  $a$ ) yang mengontrol kemiringan fungsi respons butir. Pada saat  $a$  meningkat, kemiringan fungsi respons butir juga meningkat. Kemiringan

fungsi respons butir merefleksikan seberapa baiknya kemampuan butir untuk membedakan antara individu yang memiliki nilai kemampuan  $\theta$  tinggi ataupun rendah; ini dikenal sebagai daya beda atau parameter pembeda [12]. Biasanya rentang daya beda berada antara 0-2 logit, atau nilai daya beda tidak akan melebihi +2 logit [9]. Artinya nilai  $a$  yang tinggi akan menghasilkan kurva karakteristik butir yang sangat curam, sedangkan nilai  $a$  yang rendah akan menciptakan kurva karakteristik butir yang naik secara bertahap sebagai satu fungsi kemampuan. Tingginya daya beda sebuah butir merefleksikan tingginya tingkat informasi yang diberikan oleh sebuah butir mengenai tingkat kemampuan responden. Oleh sebab itu, nilai  $a$  adalah indikator dari berapa banyak sebuah butir memberi informasi mengenai tingkat kemampuan peserta tes.

**c. Three-parameter logistic model (3-PL)**

Model 3-PL adalah model yang menambahkan parameter  $c$ , mewakili peluang menebak jawaban. Dengan adanya indeks tebakan semu (*pseudoguessing*) pada model 3-PL, memungkinkan peserta yang memiliki kemampuan rendah mempunyai peluang untuk menjawab butir soal dengan benar. Nilai  $c$  merefleksikan nilai terendah dari fungsi respon butir saat kemampuan menjadi sangat rendah (dikenal sebagai asimtot fungsi respons butir). Jadi, jika  $c=0,2$ , maka peluang menjawab benar bagi individu dengan kemampuan yang sangat rendah ialah 0,2. Karena nilai  $c$  merefleksikan hasil perilaku menebak jawaban, hal ini disebut dengan parameter *pseudo-guessing*. Indeks tebakan semu pada tes pilihan ganda terletak di sekitar seperbanyaknya pilihan jawaban. Misalnya pada tes dengan pilihan 4 jawaban, maka nilai  $c_i$  terletak di sekitar  $\frac{1}{4}$  atau 0,25 [10].

**d. Four-parameter logistic model (4-PL)**



Melalui model 4-PL akan dijelaskan bahwa siswa yang memiliki kemampuan tinggi tidak selalu menjawab soal dengan benar. Terkadang mereka mengerjakan soal dengan sembrono, sehingga soal yang seharusnya dijawab benar malah dijawab salah. Selanjutnya, dikarenakan indeks parameter butir dan kemampuan peserta merupakan hasil estimasi, maka kebenarannya bersifat probabilistik dan mengandung kesalahan pengukuran. Namun, karena bebas dari *group dependent* dan *item dependent*, maka IRT dapat digunakan untuk mengetahui karakteristik butir soal secara lebih meyakinkan dibandingkan dengan CTT [11].

## KESIMPULAN

*Item Response Theory* (IRT) merupakan salah satu bekal yang wajib dimiliki oleh calon pendidik kimia Indonesia sebagai modal daya saing terhadap lulusan pendidikan kimia dari negara-negara ASEAN. Melalui IRT, calon pendidik kimia dapat melakukan pengukuran kemampuan peserta didiknya menggunakan instrumen dengan karakteristik parameter butir yang terstandar. Instrumen terstandar dapat memberikan kesalahan pengukuran yang minimal, sehingga hasil pengukuran menjadi lebih cermat.

## UCAPAN TERIMAKASIH

1. Universitas Negeri Yogyakarta
2. Lembaga Pengelola Dana Pendidikan

## DAFTAR RUJUKAN

- [1] Agita Tarigan. 2015. *Pemerintah Harus Matangkan Pendidikan Hadapi MEA*. Diakses pada tanggal 20 Mei 2016 dari <http://www.antaraneews.com/berita/481322/pemerintah-harus-matangkan-pendidikan-hadapi-mea>

- [2] Kementerian Pendidikan dan Kebudayaan. 2007. *Permendiknas No.16, Tahun 2007, tentang Standar Kualifikasi Akademik dan Kompetensi Guru*.
- [3] Nitko, A., & Brookhart, S.M. 2011. *Educational Assessment of Students (6<sup>th</sup> Ed)*. New Jersey: Pearson Merrill Prentice Hall.
- [4] Griffin, P., & Nix, P. 1991. *Educational Assessment and Reporting: A New Approach*. Sydney: Harcourt Brace Jovanovich.
- [5] William, D. 2000. *Integrating Formative and Summative Functions of Assessment*. Makalah disajikan dalam Working Group 10 of the International Congress on Mathematics Education, di Makuhari, Tokyo.
- [6] Djemari Mardapi. 2012. *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Yogyakarta: Nuha Litera.
- [7] Djemari Mardapi. 2008. *Teknik Penyusunan Instrumen Tes dan Nontes*. Yogyakarta: Mitra Cendekia.
- [8] Ogbebor, U., & Onuka, A. 2013. Differential Item Functioning Method as an Item Bias Indicator. *Educational Research*. 4 (4). 367-373
- [9] Hambleton, R.K., & Swaminathan, H. (1985). *Items Response Theory: Principles and Application*. Boston: Kluwer-Nijhoff Publish.
- [10] Dali S. Naga. (1992). *Pengantar Teori Sekor pada Pengukuran Pendidikan*. Jakarta: Gunadarma.
- [11] Samsul Hadi. (2013). *Pengembangan Computerized Adaptive Test Berbasis Web*. Yogyakarta: Aswaja Pressindo.
- [12] Office of Assessment, Evaluation, & Research Services (OAERS). *Overview of item response theory*. Diakses pada tanggal 20 Mei 2015 dari <http://erm.uncg.edu/oaers/methodology-resources/item-response-theory/>

## TANYA JAWAB

**Penanya: Friska Septiani Silitonga**

**Pertanyaan:** Cara melihat kriteria butir dengan kemampuan siswa

**Penjawab: Kriswantoro**

---

**Jawaban:** Melihat dari hasil output analisis apakah butir tersebut mengukur kemampuan yang homogeny atau heterogen

**Penanya: Soekristin**

**Pertanyaan:** Apakah ada hak paten yang dipakai untuk acuan?

**Penjawab: Kriswantoro**

**Jawaban:** Hak paten tentu ada, semua standar dalam Item Respon Theory, banyak teori-teori dan jurnal internasional yang bisa dijadikan literature