

PENGEMBANGAN INSTRUMEN TES KEMAMPUAN BERPIKIR TINGKAT TINGGI KIMIA

Oleh:

Fatnan Asbupel¹, Damris M², Aulia Sanova³

¹Alumni Prodi Pendidikan Kimia, Jurusan PMIPA, FKIP Universitas Jambi

²Staff Pengajar Prodi Pendidikan Kimia, Jurusan PMIPA, FKIP Universitas Jambi

Program Studi Pendidikan Kimia
Jurusan Pendidikan Matematika dan Ilmu Pengetahuan Alam Fakultas Keguruan dan Ilmu
Pendidikan
Universitas Jambi
Email: asbupel96@gmail.com

ABSTRAK

Instrumen tes kemampuan berpikir tingkat tinggi (HOTS) penting untuk dapat mengukur tingkat penguasaan materi ajar peserta didik. Bagaimanapun, banyak guru yang belum atau tidak melakukan pengukuran HOTS peserta didik. Tujuan penelitian ini adalah untuk mengembangkan instrumen tes yang dapat mengukur kemampuan berpikir tingkat tinggi peserta didik pada bidang studi kimia dengan materi asam basa. Penelitian ini mengadopsi model wilson, model Oriondo dan Antonio yang dimodifikasi. Sebanyak 30 item disiapkan dan divalidasi dan diujicobakan pada peserta didik kelas XI SMAN 1 Sungai Penuh. Semua item tes yang dikembangkan fit dengan model PCM 1-PL. Tingkat kesukaran item berada pada rentang $-1,42 \leq b \leq 1,52$ dan reliabilitas yang baik dengan skor sebesar 0,98 pada tahap ujicoba. Hasil pengukuran HOTS peserta didik menunjukkan 57,14% pada HOTS sangat tinggi dan sisanya 42,86% HOTS sangat rendah. Instrumen tes yang dikembangkan untuk mengukur HOTS peserta didik dengan materi asam basa memiliki tingkat kesukaran dan reliabilitas yang baik.

Kata kunci : Pengembangan tes, Penilaian, Kemampuan Berpikir Tingkat Tinggi, HOTS, Kimia, SMA

Pendahuluan

Evaluasi merupakan faktor penting didalam pendidikan. Melalui evaluasi diperoleh informasi berupa data dari hasil pengukuran sehingga ketercapaian tujuan pendidikan dapat diketahui. Sebagaimana diatur dalam Undang-Undang RI Nomor 20 Tahun 2003 tentang sistem pendidikan nasional. Evaluasi pendidikan adalah kegiatan pengendalian, penjaminan, dan penetapan mutu pendidikan terhadap berbagai komponen pendidikan pada setiap jalur, jenjang, dan jenis pendidikan sebagai pertanggungjawaban penyelenggaraan pendidikan. Memperbaiki kualitas pendidikan sebaiknya dilakukan dengan sistem evaluasi yang lebih baik. Sistem evaluasi (kegiatan pengukuran, pengujian, penilaian) ini, selain prosedurnya yang harus sistematis, pelaksanaannya pun harus memiliki akuntabilitas yang tinggi, serta hasilnya diharapkan mampu dijadikan tolak ukur dari tujuan yang diharapkan.

Pengukuran dan penilaian merupakan dua hal yang tidak terlepas dari sistem evaluasi. Pengukuran merupakan proses pemberian angka dengan membandingkan hasil amatan dengan kriteria yang telah ditetapkan. Hal ini sejalan dengan pendapat Olivo, *et al* (2014) yang menyatakan bahwa berdasarkan hasil pengukuran maka diperoleh informasi sebagai bahan masukan kepada penyelenggara pendidikan dalam pengambilan keputusan. Proses pengukuran memerlukan alat ukur yang baik, sebab alat ukur ini yang nantinya akan memberikan informasi mengenai gambaran seseorang yang sedang diukur. Namun dalam melakukan pengukuran, alat ukur yang digunakan harus lah valid dan reliabel agar menghasilkan data dan memberikan informasi yang akurat. Data yang akan dihasilkan nantinya dijadikan acuan dalam melakukan penilaian.

Alat ukur yang digunakan oleh sekolah salah satunya adalah dengan menggunakan instrumen tes. Instrumen tes merupakan alat ukur yang terdiri dari butir-butir pertanyaan untuk mengukur kemampuan peserta didik dalam mencapai

tujuan pembelajaran yang telah dilakukan. Ketepatan dalam pelaksanaan penilaian ditentukan oleh baik buruknya perangkat tes yang digunakan oleh guru. Jika sebuah perangkat tes memiliki karakteristik yang tidak baik maka hasil penilaian menjadi tidak valid, oleh sebab itu diperlukan kemampuan guru dalam membuat instrumen penilaian yang berkualitas. Hal ini diungkapkan pula oleh Nurung (2008) dalam penelitiannya yang mengatakan bahwa pengelolaan ujian dan mutu bahan ujian yang digunakan perlu mendapat perhatian agar hasil tes dapat mencerminkan kemampuan peserta didik yang sebenarnya. Rumusan penilaian yang diperoleh dari hasil pengukuran berdasarkan perangkat tes yang digunakan berpengaruh pada berbagai keputusan atau kebijakan pendidikan. Sehingga alangkah baiknya perangkat tes yang digunakan untuk mengukur hasil belajar peserta didik memiliki kualitas yang baik dan berdasarkan kriteria dari berpikir tingkat tinggi yang digunakan.

Menurut Brookhart (2010) kemampuan berpikir tingkat tinggi (HOTS) adalah (1) berpikir tingkat tinggi berada pada bagian atas Taksonomi kognitif Bloom, (2) tujuan pengajaran di balik Taksonomi kognitif yang dapat membekali peserta didik untuk melakukan transfer pengetahuan, (3) mampu berpikir artinya peserta didik mampu menerapkan pengetahuan dan keterampilan yang mereka kembangkan selama belajar pada konteks yang baru. Dalam hal ini yang dimaksud “baru” adalah aplikasi konsep yang belum terpikirkan sebelumnya oleh peserta didik, namun konsep tersebut sudah diajarkan, ini berarti belum tentu sesuatu yang universal baru. Berpikir tingkat tinggi berarti kemampuan peserta didik untuk menghubungkan pembelajaran dengan hal-hal lain yang belum pernah diajarkan.

Kemampuan berpikir tingkat tinggi merupakan aspek penting dalam proses belajar mengajar. Kemampuan berpikir merupakan hal yang mendasar dalam proses pendidikan. Pikiran seseorang dapat mempengaruhi kemampuan belajar, kecepatan dan efektivitas pembelajaran.

Kemampuan berpikir sangat berkaitan dengan proses pembelajaran. Peserta didik yang dilatih untuk berpikir menunjukkan dampak positif pada perkembangan pendidikan mereka. Dengan demikian, guru bersama peneliti merencanakan suatu usaha yaitu memperbaiki cara berpikir peserta didik dari tingkat rendah ke berpikir tingkat tinggi, dengan melihat pengembangan instrumen tes yang telah teruji karakteristik butirnya.

Kimia merupakan ilmu yang termasuk rumpun IPA, oleh karenanya kimia mempunyai karakteristik sama dengan IPA. Karakteristik tersebut adalah objek ilmu kimia, cara memperoleh, serta kegunaannya (BSNP, 2006). Asam basa merupakan bagian dari ilmu kimia yang diajarkan di SMA. Pada materi asam basa banyak mengajarkan tentang fakta, teori, prinsip dan pengetahuan kimia dalam kehidupan sehari-hari. Setelah dilakukan analisa kurikulum dan silabus untuk materi asam basa serta pengamatan langsung proses pembelajaran di SMA Negeri Kota Sungai Penuh, untuk materi larutan asam basa layak digunakan untuk dikembangkan menjadi instrumen tes berpikir tingkat tinggi di kelas XI SMA Negeri Kota Sungai Penuh.

Berdasarkan hasil survey dan wawancara yang telah dilakukan oleh peneliti di Sekolah Menengah Atas (SMA) Negeri di Kota Sungai Penuh bahwasanya selama ini guru di sekolah tersebut khususnya guru mata pelajaran kimia, dalam membuat instrumen tes guru hanya membuat soal tanpa melihat karakteristik dari instrumen tes tersebut. Guru hanya membuat, mengujikan dan mengambil jawaban dari peserta didik, tanpa melihat layak atau tidak layaknya instrumen tes tersebut digunakan. Selain itu juga, kebanyakan para guru membuat instrumen tes tersebut hanya pada tingkat kognitif C1 (mengingat), C2 (memahami), dan C3 (mengaplikasikan) menurut Taksonomi Bloom, sehingga peserta didik hanya terbiasa mengerjakan instrumen tes pada kategori berpikir tingkat rendah. Seharusnya guru sudah terbiasa membuat instrumen tes pada tingkat kognitif C4

(menganalisis), C5 (mengevaluasi), dan C6 (mencipta) menurut Taksonomi Bloom, sehingga peserta didik sudah terbiasa mengerjakan instrumen tes pada kategori tingkat tinggi.

Kajian Pustaka

Menurut Taksonomi Bloom yang telah direvisi proses kognitif terbagi menjadi kemampuan berpikir tingkat rendah (*Lower Order Thinking*) dan kemampuan berpikir tingkat tinggi (*Higher Order Thinking*). Kemampuan yang termasuk LOT adalah kemampuan mengingat (*remember*), memahami (*understand*), dan menerapkan (*apply*), sedangkan HOT meliputi kemampuan menganalisis (*analyze*), mengevaluasi (*evaluate*), dan menciptakan (*create*) (Anderson & Krathwohl, 2010). Taksonomi Bloom sudah lama diterapkan dalam bidang pendidikan dan sudah lama digunakan. Taksonomi Bloom masih digunakan dalam banyak kurikulum dan bahan pengajaran (Brookhart, 2010). Dengan demikian, kemampuan berpikir tingkat tinggi kimia meliputi kemampuan kimia dalam menganalisis, mengevaluasi, dan menciptakan.

Maka dari itu, melalui pembelajaran kimia diharapkan peserta didik dapat mengembangkan diri dalam berpikir. Peserta didik dituntut tidak hanya memiliki kemampuan berpikir tingkat rendah, tetapi sampai pada kemampuan berpikir tingkat tinggi.

Ada dua teori pengukuran pendidikan yang saat ini berkembang dan banyak digunakan dalam merancang dan menganalisis suatu tes. Teori yang pertama adalah teori tes klasik yang dikembangkan sejak tahun 1940 dan telah digunakan secara luas, kedua adalah teori respons butir yang menggunakan lebih banyak asumsi dibandingkan teori pertama. Teori tes klasik telah banyak berjasa dalam bidang pengukuran dan bahkan masih digunakan sampai sekarang. Namun demikian, oleh karena teori tes klasik memiliki berbagai keterbatasan dengan adanya sifat group dependent dan item dependent maka

munculnya teori respons butir menjadi sangat berguna dan terus dikembangkan karena mampu mengatasi keterbatasan tersebut (Hambleton, *et al.* 1991).

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) (Dali S. Naga: 1992). 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 (Samsul Hadi: 2013).

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 responsnya 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.

Metode Penelitian

Penelitian ini termasuk jenis penelitian pengembangan dengan pendekatan kuantitatif. Penelitian pengembangan instrumen ini menggunakan modifikasi Model Wilson, Model Oriundo dan Antonio (1998). Langkah-langkah pengembangan instrumen tes ini adalah : (1) perancangan tes, (2) uji coba tes, dan (3) pengukuran tes. Alasan kenapa memilih model tersebut karena model Wilson Oriundo dan Antonio lebih efektif dan sesuai dengan prosedur pengembangan instrumen tes.

Tahap perancangan tes meliputi: (1) penentuan tujuan tes, (2) penentuan kompetensi yang akan diujikan, (3) penentuan materi yang akan diujikan, (4) penyusunan kisi-kisi tes, (5) penulisan item berdasarkan prinsip-prinsip pengembangan berpikir tingkat tinggi, (6) validitas item, dan (7) perbaikan item dan perakitan tes. Adapun tahap uji coba tes meliputi: (1) melakukan uji coba tes, (2) menganalisis butir tes, dan (3) memperbaiki butir tes. Tahap terakhir dalam pengembangan tes ini adalah pengukuran tes meliputi: (1) merakit tes, (2) melaksanakan tes, dan (3) menafsirkan hasil tes.

Uji coba dilaksanakan di SMA Negeri 4 Kota Sungai Penuh dan pengukuran dilaksanakan di SMA Negeri 1 Kota Sungai Penuh. Ujicoba yang dilakukan bertujuan untuk menyempurnakan instrumen tes dengan mengujikan di lapangan peserta didik SMA kelas XI semester genap dilihat dari tes hasil belajar pada materi asam basa, titrasi asam basa, dan larutan penyangga pada semester genap tahun ajaran 2016/2017.

Subjek penelitian ini peserta didik kelas XI SMA Negeri di Kota Sungai Penuh. Banyaknya sampel yang akan diambil dari populasi menggunakan acuan Tabel Morgan dengan derajat kepercayaan 95% (Krejcie & Morgan, 1970). Berdasarkan tabel tersebut, jika ukuran populasi 2000 peserta didik, maka jumlah minimum sampel yang harus ditarik adalah 322 peserta didik. Sementara itu, besarnya ukuran sampel minimal untuk

model 1-PL (Rasch model) adalah 200 (Wright & Stone, 1979) atau 150-250 (Linacre, 2012). Berdasarkan pertimbangan tersebut maka dalam penelitian ini ditetapkan sampel sejumlah sedikitnya 150 untuk uji coba dan 150 peserta didik untuk pengukuran.

Instrumen pengumpulan data penelitian ini adalah berpikir tingkat tinggi kimia yang merupakan perangkat tes. Perangkat tes ini berupa soal pilihan ganda beralasan yang masing-masing terdapat 5 pilihan jawaban dan alasan. Berkenaan dengan instrumen pengumpulan data berupa perangkat tes, maka teknik pengumpulan data tentu juga tes. Semua subjek uji coba diminta mengerjakan tes sesuai dengan perangkat tes yang diterima, yakni : perangkat tes. Lembar jawaban peserta didik dikumpulkan setelah pelaksanaan uji coba.

Analisis data dimulai dengan mendeskripsikan kelayakan (karakteristik) instrumen tes Berpikir Tingkat Tinggi Kimia kelas XI menggunakan *Item Response Theory* dengan bantuan program *Winsteps* dan *QUEST*. Program ini dipilih karena memiliki kelebihan yakni : (1) dapat menganalisis data dikotomis dan politomis beserta kombinasinya, (2) tersedianya hasil analisis baik berdasarkan teori tes klasik maupun teori respon modern, dan (3) hasil analisis teori respon modern didasarkan pada model likelihood maximum menggunakan model logistik satu parameter (Subali & Suyata, 2011).

Analisis berdasarkan pendekatan IRT dapat dilakukan dengan menguji asumsi unidimensi melalui analisis kesesuaian (fit) atau analisis faktor eksploratori. Butir tes bersifat unidimensi berarti bahwa butir tersebut mengukur satu kemampuan. Jika asumsi unidimensi telah terpenuhi, maka secara otomatis asumsi independensi lokal juga telah terpenuhi. Salah satu indikasi apakah butir tes bersifat unidimensi adalah data sesuai (fit) dengan model Rasch. Untuk mengetahui apakah model Rasch dapat memprediksi respons setiap responden maka digunakan statistik *infit mean-square* (IMS) dan *outfit mean-square* (OMS). Statistik

IMS dan OMS merupakan ukuran tingkat kesesuaian antara data observasi dan nilai yang diprediksi oleh model. Butir tes dikatakan fit model jika memiliki nilai IMS dan OMS berkisar dari 0,5 sampai 1,5 (Linacre, 2012).

Tabel 1. Rentang Nilai IMS dan OMS

Nilai	Implikasi bagi pengukuran
>2,0	Merusak sistem pengukuran
1,5 – 2,0	Tidak mempunyai makna bagi pengukuran
0,5 – 1,5	Bermanfaat bagi pengukuran
<0,5	Tidak bermanfaat bagi pengukuran, tetapi tidak merusak

Sementara itu, karakteristik pada parameter butir berupa tingkat kesulitan butir diinterpretasi menggunakan kriteria dari Baker (2001) berikut :

Sangat mudah	Mudah	Sedang	Sukar	Sangat sukar
-2,0	-0,5	0	+0,5	+2,0

Tes memberikan makna setelah dianalisis hasilnya. Hasil kerja peserta didik selanjutnya dianalisis agar diperoleh data kemampuan berpikir tingkat tinggi kimia peserta didik pada kelas XI SMA. Secara garis besar untuk menganalisis tes menggunakan teori respon butir (b). Pada model yang digunakan IRT analisis tes dilakukan berdasarkan butir, banyak peserta tes, dan model yang digunakan untuk analisis adalah IRT dengan 1-PL. Penskoran untuk tes pilihan ganda beralasan ini dengan politomis menurut PCM.

Hasil pengukuran disajikan dalam bentuk : (1) grafik distribusi frekuensi kemampuan (*ability*), dan (2) chart persentase berdasarkan kategori level kemampuan. Untuk mengetahui tingkat Berpikir Tingkat Tinggi tersebut digunakan kategori berdasarkan rata-rata ideal dan simpangan baku ideal. Hal ini diterapkan dengan asumsi bahwa kemampuan berpikir tingkat tinggi kimia berdistribusi normal. Penentuan skor rata-rata ideal (M_i) dan skor simpangan baku ideal (SB) didasarkan pada

skor tertinggi dan terendah dari variabel penelitian yang dinyatakan dengan Azwar (2002).

Dengan :

Mi : rerata ideal

Sbi : simpangan baku ideal

Xmak: skor tertinggi

Xmin: skor terendah

Tabel 2 Interval Nilai Level Kemampuan

No	Interval Kemampuan	Level
1	$Mi + 1,5 SBi < \theta$	Sangat Tinggi
2	$Mi + 0,5 SBi < \theta \leq Mi + 1,5 Sbi$	Tinggi
3	$Mi - 0,5 SBi < \theta \leq Mi + 0,5 Sbi$	Sedang
4	$Mi - 1,5 SBi < \theta \leq Mi - 0,5 Sbi$	Rendah
5	$0 < Mi - 1,5 Sbi$	Sangat Rendah

Interpretasi hasil pengukuran dilakukan atas dasar skor secara politomus dengan 4 kategori dari hasil pengukuran. Berdasarkan hasil analisis dengan model *Rasch* program *Winsteps* diperoleh kemampuan (*ability*).

Berdasarkan koefisien tingkat kesulitan butir, ICC, fungsi informasi dan SEM dapat dideskripsikan karakteristik tes. Atas dasar koefisien reabilitas dan tingkat kesulitan butir dapat dinyatakan kelayakannya tes digunakan sebagai dasar dalam pengambilan keputusan terhadap peserta didik.

Hasil dan Pembahasan

Hasil Pengembangan Tes

Pada tahap perancangan tes ada beberapa kegiatan yang dilakukan sehingga diperoleh: (1) penentuan tujuan tes, (2) penentuan kompetensi, (3) penentuan materi, (4) penyusunan kisi-kisi tes, (5) penulisan item, dan terakhir yaitu (6) validasi dari tim ahli. Instrumen tes yang dihasilkan terlebih dahulu dilakukan validasi secara kualitatif dan kuantitatif. Validasi dilakukan oleh ahli pendidikan kimia. Validator memberikan penilaian pada lembar instrumen tes kemampuan berpikir tingkat tinggi kimia. Lembar penilaian

instrumen merupakan salah satu bagian dari validitas isi untuk membangun tes. Dari hasil analisis terdapat 30 butir yang dinyatakan valid yaitu $V > 0,956$. Butir dinyatakan valid jika memenuhi kriteria $V > 0,73$ dengan kategori penilaian sebanyak 4 skala (Aiken, 1980).

Hasil Ujicoba Produk

Pada kegiatan ujicoba ini, instrumens tes berpikir tingkat tinggi kimia, diujicobakan di sekolah yaitu SMA Negeri 1 kota Sungai Penuh. Ujicoba melibatkan 174 peserta didik. Setelah kegiatan ujicoba dilaksanakan, jawaban peserta tes diskor kemudian dianalisis. Analisis ujicoba dilakukan agar diperoleh karakteristik instrumen yang baik. Dalam analisis ini dilakukan model PCM 1-PL yang melihat tingkat kesukaran butir soal. Sebelum melihat tingkat kesukaran tersebut, dalam teori respon butir harus memenuhi syarat (asumsi) dan tahapan berupa unidimensi, kecocokan butir dan tingkat kesukaran butir. Kegiatan analisis karakteristik instrumen ini dilakukan dengan bantuan program excel dan Quest.

Uji Unidimensi

Uji asumsi unidimensi dilakukan untuk mengetahui apakah instrumen tes yang digunakan mengukur satu macam ciri (trait). Asumsi unidimensi dibuktikan menggunakan analisis faktor dan hasil analisis empirisnya. Sebelum melakukan pembuktian terhadap asumsi unidimensi, terlebih dahulu dilihat kelayakan sampel uji yang digunakan melalui KMO-MSA dan *Bartlett's Test of Sphericity*.

Tabel 3 KMO-MSA dan *Bartlett's Test of Sphericity*

KMO and Bartlett's Test		
<i>Kaiser-Meyer-Olkin Measure of Sampling Adequacy</i>		0.636
<i>Bartlett's Test of Sphericity</i>	<i>Approx. Chi-Square</i>	3690.169
	<i>Df</i>	435
	<i>Sig.</i>	.000

Kaiser Meyer Olkin Measure of Sampling Adequacy (KMO-MSA) adalah

indeks perbandingan jarak antara koefisien korelasi dengan koefisien korelasi parsialnya. Jika jumlah kuadrat koefisien korelasi parsial diantara seluruh pasangan variabel bernilai kecil jika dibandingkan dengan jumlah kuadrat koefisien korelasi, maka akan menghasilkan nilai KMO mendekati 1. Nilai KMO-MSA dianggap mencukupi jika lebih dari 0,5 (Field, 2009). Hasil penelitian menunjukkan bahwa nilai KMO-MSA sebesar 0,636. Menurut Kaiser, Meyer, & Olkin nilai KMO-MSA sebesar 0,636 termasuk kategori moderat karena berada dalam rentang 0,6-0,69.

Penyimpulan unidimensionalitas secara grafis menggunakan *scatter plot*. Menurut Jacoby (2012), dimensionalitas dapat diketahui dengan cara “*Look for an ‘elbow’ in the scree plot. Dimensionality corresponds to the number of dimensions that falls just prior to the elbow*”. Pendapat tersebut serupa dengan pendapat Hambleton & Rovinelli (1986) yang menyatakan bahwa biasanya jumlah faktor yang signifikan ditentukan dengan munculnya sebuah “siku dalam plot”, jumlah eigen di sebelah kiri siku ini diinterpretasikan sebagai banyaknya dimensi yang terbentuk. *Scree plot* pada Gambar 1 menunjukkan bahwa telah terbentuk siku, dengan satu titik di sebelah kiri siku. Dengan mengacu kepada pendapat para ahli di atas, maka dapat disimpulkan bahwa asumsi unidimensi terpenuhi.

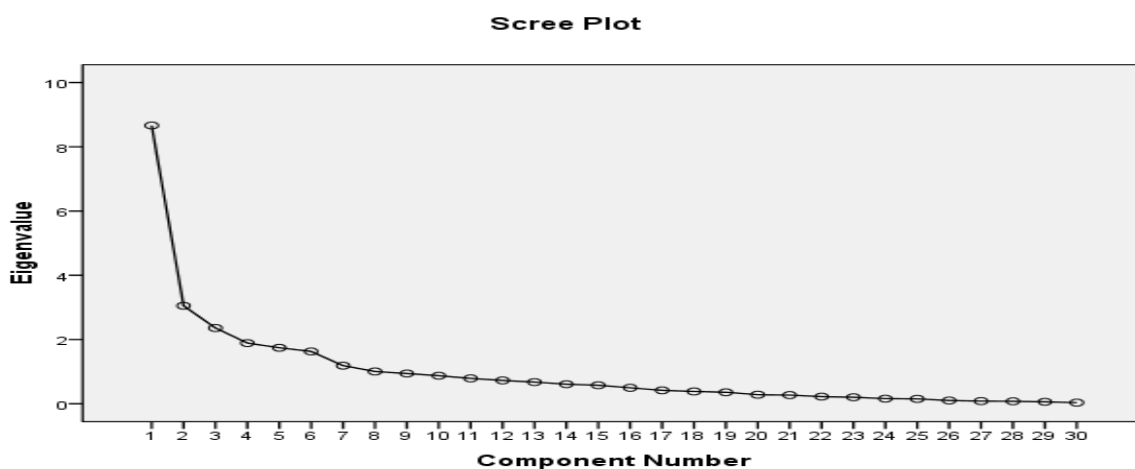
Kecocokan Butir

Pengujian penetapan fit setiap item terhadap model mengikuti kaidah Linacre (2012), statistik IMS dan OMS merupakan ukuran tingkat kesesuaian antara data observasi dan nilai yang diprediksi oleh model. Butir tes dikatakan fit model jika memiliki nilai IMS dan OMS berkisar dari 0,5 sampai 1,5. Item bisa dikatakan fit dengan model apabila salah satu dari keduanya memenuhi asumsi item tersebut sudah bisa dikatakan fit dengan model. Hasil analisis dengan menggunakan program *Quest* didapatkan hasil sebagai berikut.

Dari Tabel 4.2 di atas dapat disimpulkan bahwa, dari 30 item pada instrumen ini semuanya cocok dengan model. Lebih jelas bisa dilihat di Gambar 4.2 map kecocokan item terhadap model dengan bantuan program *Quest*.

Tingkat Kesukaran Butir

Perhitungan tingkat kesukaran tes dilakukan dengan menggunakan program aplikasi *Quest*. Tingkat kesukaran butir dapat kita ketahui dari butir yang paling sulit dan butir yang paling mudah. Apabila tingkat kesukaran bernilai positif maka butir tersebut bisa dikatakan sulit dan apabila tingkat kesukaran bernilai negatif butir dikatakan mudah. Rentang tingkat kesukaran dimulai dari rentang -0,88 sampai dengan 1,04.



Gambar 1. Scree Plot

Nilai yang semakin mendekati -2 logit mengindikasikan butir semakin mudah, dan nilai yang mendekati +2 logit mengindikasikan butir semakin sukar. Tingkat kesukaran yang telah dirumuskan oleh peneliti memang tidak sesuai dengan tingkat kesukaran hasil empirik. Hal ini dikarenakan dalam membuat item tersebut, peneliti mengklasifikasikan item ke dalam tingkat kesukaran tertentu (mudah, sedang, dan sukar) hanya berdasarkan intuisinya (Stanley & Wang, 1970). Belum tentu item yang dianggap peneliti sebagai item”sulit” juga dirasakan sulit oleh peserta didik karena sangat sulit menentukan seberapa sulit item dalam suatu tes sebelum peserta didik melakukan tes (Baker, 2001). Untuk lebih jelas bisa dilihat pada gambar 2 map kecocokan item terhadap model.

Reliabilitas

Reabilitas atau keandalan merupakan koefisien yang menunjukkan tingkat keajegan atau konsistensi hasil pengukuran suatu tes. Reliabilitas merupakan penerjemahan dari kata reliability yang mempunyai kata rely dan ability. Pengukuran yang memiliki reliabilitas tinggi disebut

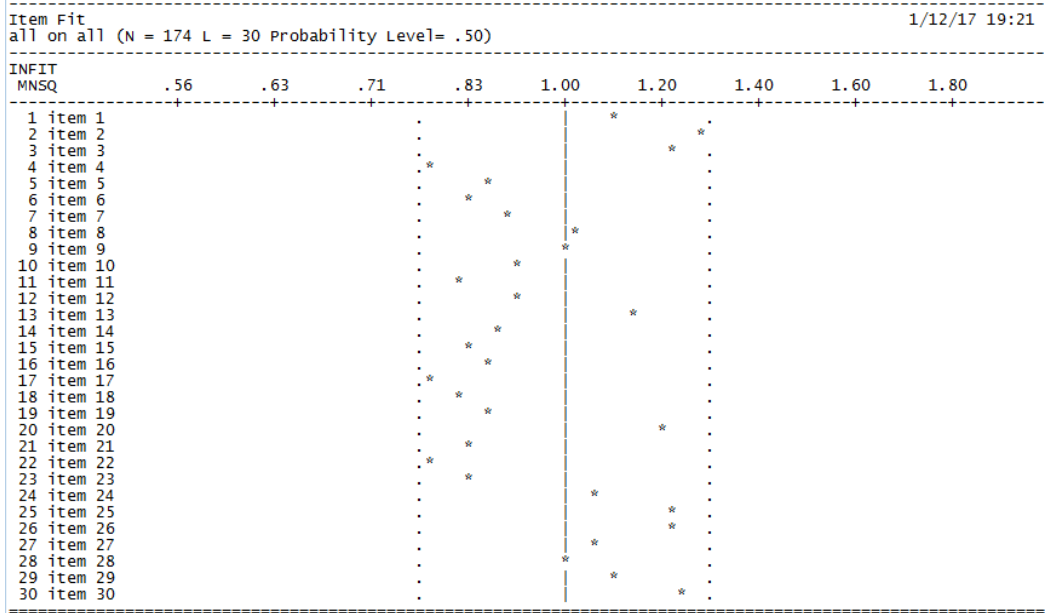
sebagai pengukuran yang reliabel (reliable). Ide pokok yang terkandung dalam konsep reliabilitas adalah sejauhmana hasil suatu pengukuran dapat dipercaya (Azwar, 2009). Pada pengujian reliabilitas diperoleh hasil sebesar 0,98. Lebih jelasnya dapat dilihat pada gambar 3 hasil analisis reliabilitas.

SUMMARY OF 30 MEASURED Item									
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	
MEAN	424.4	175.0	.00	.08	1.04	-.7	1.06	-.5	
S.D.	111.8	.0	.69	.02	.49	5.2	.54	5.0	
MAX.	649.0	175.0	1.68	.15	2.14	9.9	2.67	9.9	
MIN.	208.0	175.0	-1.51	.07	.21	-9.9	.21	-9.9	
REAL RMSE	.09	TRUE SD	.68	SEPARATION	7.25	Item	RELIABILITY	.98	
MODEL RMSE	.08	TRUE SD	.68	SEPARATION	8.17	Item	RELIABILITY	.99	
S.E. OF Item MEAN	= .13								

Gambar 3. Hasil Analisis Reliabilitas

Azwar (2008) mengemukakan bahwa reliabilitas dinyatakan oleh koefisien reliabilitas yang angkanya berada dalam rentang dari 0 sampai dengan 1.00. semakin tinggi koefisien reliabilitas mendekati angka 1.00 berarti semakin tinggi reliabilitasnya.

Sebaliknya, koefisien yang semakin rendah mendekati angka 0 berarti semakin rendah reliabilitasnya. Jadi dapat disimpulkan bahwa reliabilitas pada instrumen tes kemampuan berpikir tingkat tinggi kimia dalam kategori baik yaitu 0,98.



Gambar 2. Map Kecocokan Item Terhadap Model

Kajian Produk Akhir

Data hasil pengukuran berupa kemampuan yang dijadikan sumber informasi tentang gambaran kemampuan berpikir tingkat tinggi kimia siswa SMA Negeri 4 di kota Sungai Penuh diperoleh melalui penafsiran skor. Penafsiran skor menghasilkan nilai. Nilai tersebut dapat disajikan dalam bentuk bilangan atau dalam bentuk predikat. Hasil penafsiran skor atau hasil penilaian dapat dimanfaatkan untuk berbagai kepentingan antara lain untuk perbaikan kualitas pembelajaran, pelaporan hasil belajar dan sebagainya. Pengukuran dilaksanakan di SMA Negeri 4 Sungai Penuh.

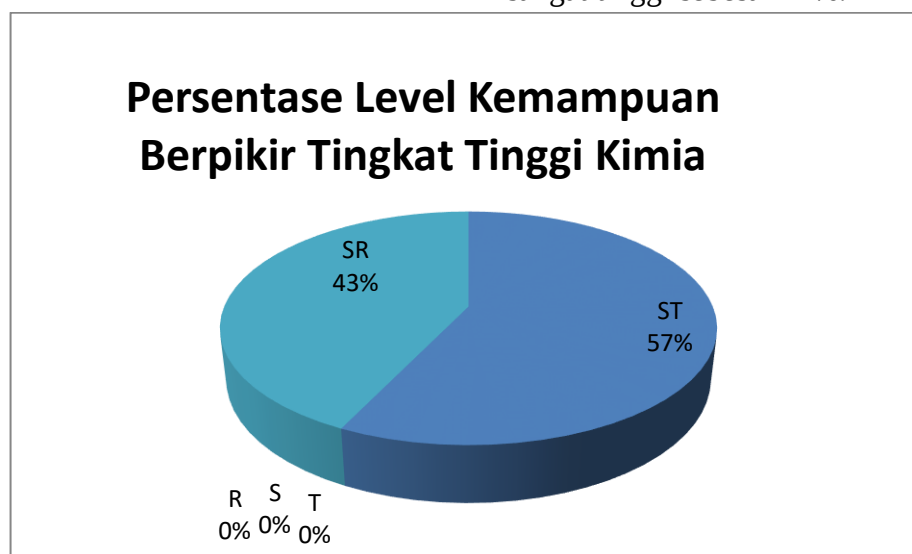
Hasil penilaian kemampuan berpikir tingkat tinggi juga dapat disajikan dalam bentuk predikat sangat rendah sampai sangat tinggi. Untuk menghasilkan nilai dalam bentuk predikat tersebut dapat dilakukan dengan cara membuat kategori skor. Tes kemampuan berpikir tingkat tinggi dalam penelitian ini memiliki teta maksimal 0.83 dan minimum -0.94. Dengan demikian rentangan skor sebesar 2,03 dan nilai rata-rata ideal sebesar -0.055. Berdasarkan rata-rata ideal dan simpangan baku ideal tersebut kategori kemampuan berpikir tingkat tinggi adalah sebagai berikut:

Tabel 4. Interval Nilai Level Kemampuan

No.	Interval Kemampuan	Level
1	$-0,0825 < \theta$	Sangat Tinggi
2	$-0,0641 < \theta \leq -0,0825$	Tinggi
3	$-0,0458 < \theta \leq -0,0641$	Sedang
4	$-0,0275 < \theta \leq -0,0458$	Rendah
5	$0 < -0,0275$	Sangat Rendah

Hasil penilaian kemampuan berpikir tingkat tinggi tersebut terlihat bahwa peserta tes yang memiliki kemampuan sangat tinggi sebesar 57,14 %, peserta tes memiliki kemampuan tinggi sebanyak 0%, peserta tes memiliki kemampuan sedang sebanyak 0%, peserta tes memiliki kemampuan rendah sebanyak 0%, dan peserta tes memiliki kemampuan sangat rendah sebesar 42,86 %. Persentase tersebut menunjukkan banyaknya peserta tes yang memiliki yang lebih dari skor rata-rata ideal, ini menunjukkan banyaknya peserta tes yang memiliki diatas rata-rata.

Dari Gambar 4 dapat kita jabarkan bahwa, gambaran kemampuan berpikir tingkat tinggi kimia peserta didik di SMA Negeri 4 Sungai Penuh untuk kategori level sangat tinggi sebesar 57%, kemampuan tinggi sebesar 0%, kemampuan sedang sebesar 0%, kemampuan rendah sebesar 0%, dan kemampuan sangat rendah sebesar 43%. Dari uraian diatas gambaran kemampuan berpikir tingkat tinggi hampir seimbang dengan selisih antara kategori sangat rendah dan sangat tinggi sebesar 14%.



Gambar 4. Persentase level kemampuan

Simpulan dan Saran

Berdasarkan tujuan penelitian pengembangan yang telah dilakukan, dapat ditarik kesimpulan sebagai berikut:

1. Instrumen kemampuan berpikir tingkat tinggi kimia dikembangkan pada kemampuan menganalisis, mengevaluasi, dan menciptakan untuk materi larutan asam basa, stoikiometri reaksi dan titrasi, serta larutan penyangga yang terdiri dari 30 item soal pilihan ganda beralasan.
2. Instrumen kemampuan berpikir tingkat tinggi kimia memiliki karakteristik sebagai instrumen yang memenuhi syarat digunakan untuk mengukur, yakni:
 - a. Instrumen kemampuan berpikir tingkat tinggi kimia telah memenuhi validitas isi dengan expert judgment dan telah mendapatkan bukti empiris fit dengan *Partial Credit Model* (PCM) berdasarkan data politomus empat kategori.
 - b. Seluruh item pada Instrumen kemampuan berpikir tingkat tinggi dalam kriteria baik.
 - c. Reliabilitas Instrumen kemampuan berpikir tingkat tinggi kimia telah memenuhi syarat yang tergolong kategori tinggi.
 - d. Berdasarkan fungsi informasi, Instrumen kemampuan berpikir tingkat tinggi kimia sangat tepat digunakan untuk mengukur kemampuan berpikir tingkat tinggi kimia peserta didik.
3. Kemampuan berpikir tingkat tinggi kimia di SMA Negeri Kota Sungai Penuh dominan berada pada level sangat tinggi sebesar 57%, dan kemampuan sangat rendah sebesar 43%. Dari uraian gambaran menunjukkan bahwa instrumen tes berada pada level kemampuan berpikir tingkat tinggi.

Berdasarkan simpulan dan keterbatasan penelitian ini, dapat diberikan saran pemanfaatan produk sebagai berikut:

1. Dalam proses pembelajaran guru seharusnya lebih menekankan pada aspek menganalisis dan mencipta terbukti

dengan didapatkan hasil yang diperoleh pada penelitian ini.

2. Bagi guru-guru yang akan melihat kemampuan berpikir tingkat tinggi peserta didik dalam mata pelajaran kimia, agar menggunakan instrumen yang dihasilkan dalam penelitian ini, dikarenakan instrumen tersebut sudah teruji karakteristiknya.
3. Bagi guru-guru yang ingin mengembangkan sendiri instrumen kemampuan berpikir tingkat tinggi kimia agar mengikuti langkah-langkah pengembangan instrumen yang digunakan dalam penelitian ini agar diperoleh instrumen yang valid.

Daftar Pustaka

- Aiken, L. W. 1980. *Content validity and reliability of single items or questionnaires*. Malibu: Pepperdine University.
- Anderson, L. W. & Krathwohl, D. R. 2010. *Kerangka Landasan Untuk Pembelajaran, Pengajaran dan Asesmen; Revisi Taksonomi Pendidikan Bloom* (cetakan I). Yogyakarta: Pustaka Pelajar.
- Azwar, S. 2002. *Tes prestasi: fungsi dan pengembangan pengukuran prestasi belajar*. Yogyakarta: Pustaka Pelajar.
- Azwar, S. 2008. *Reliabilitas dan validitas*. Yogyakarta: Pustaka Pelajar.
- Azwar, S. 2009. *Sikap Manusia, Teori dan Pengukurannya*. Jakarta: Pustaka Pelajar.
- Baker, F. B. 2001. *The basics of item response theory* (2nd ed). Wisconsin: ERIC Clearinghouse on Assessment and Evaluation.
- Brookhart, S. M. 2010. *How to asses higher order thinking skills in your classroom*. Alexandria: ASCD.

- BSNP. 2006. *Standar kompetensi dan kompetensi dasar mata pelajaran kimia untuk SMA dan MA*. Jakarta: BSNP-Depdiknas.
- Dali S. Naga. 1992. *Pengantar Teori Sekor pada Pengukuran Pendidikan*. Jakarta: Gunadarma.
- Field, A. 2009. *Discovering statistics using SPSS (3rd ed.)*. London: Sage Publication Ltd.
- Hambleton, R.K., & Rovinelli, R. J. 1986. Assessing the Dimensionality of a Set of Test Items. *Applied Psychological Measurement*, 10(3), 287-302.
- Hambleton, R.K., Swaminathan, H., & Rogers, H.J. 1991. *Fundamental of item response theory*. London: Sage Publication.
- Jacoby, W.G. 2012. *Multidimensional scaling: an introduction*. Paper presented at the Workshop in Methods Indiana University. Bloomington, USA.
- Krejcie, R. V., & Morgan, D. W. 1970. *Determining sampel size for research activities. Educational and Psychological Measurement*. 30. 607-610.
- Linacre, J. M. 2012. *A user's guide to winsteps-ministeps rasch model computer program (program manual 3.75.0)*. Chicago.
- Nurung, M. 2008. *Kualitas tes ujian akhir sekolah berstandar nasional (UASBN) IPA SD tahun pelajaran 2007/2008 di kota kendari. Tesis magister, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta*.
- Olivo, S. A., Cumming, G. G., Fuentes. J., Saltaji, H., et al. 2014. Identifying items to assess methodological quality in physical therapy trials: a factor analysis. *Journal Physical Theraphy*, 94, 1227-1284.
- Oriondo, L.L. & Dallo-Antonio, E.M. 1998. *Evaluation educational outcomes*. Manila : Rex Printing Compagny, inc.
- Samsul Hadi. (2013). *Pengembangan Computerized Adaptive Test Berbasis Web*. Yogyakarta: Aswaja Pressindo.
- Stanley, J.C., & Wang, M.D. 1968. *Differential weighting: a survey of methods and empiricial studies*. New York: Departement of Health, Education, &Walfare.
- Subali, B. & Suyata, P. 2011. *Panduan analisis data pengukuran pendidikan untuk memperoleh bukti empirik kesahihan menggunakan program quest*. Yogyakarta: LPPM UNY.
- Wright, B. D & Stone, M. H. 1979. *Best test design*. Chicago: MESA Press.