

PENGEMBANGAN INSTRUMEN DIAGNOSIS KESULITAN BELAJAR PADA PEMBELAJARAN KIMIA DI SMA

¹⁾Sri Yamtinah, ²⁾Budiyono

¹⁾Program Studi Pendidikan Kimia FKIP UNS, ²⁾PPs Universitas Sebelas Maret

¹⁾jengtina_sp@yahoo.com, ²⁾bud@uns.ac.id

Abstrak

Tujuan penelitian ini untuk: (1) mengembangkan instrumen *ordered multiple-choice* untuk diagnosis kesulitan belajar Stoikiometri kelas X, (2) menetapkan karakteristik dari instrumen, dan (3) membuat profil diagnostik peserta didik sebagai laporan yang informatif. Penelitian pengembangan ini menggunakan model Borg & Gall. *Preliminary study* menghasilkan pokok bahasan tersulit Stoikiometri. Penetapan atribut dan hierarki atribut dengan *focus group discussion* dilanjutkan dua putaran teknik Delphi. Indikator dan rumusan butir soal ditelaah *expert* untuk membuktikan validasi isi dengan rumus *Aiken*. Ujicoba terbatas dan ujicoba menengah pada sekolah kategori tinggi, sedang dan rendah. Ujicoba kelayakan dilakukan pada sekolah-sekolah di wilayah Surakarta, Karanganyar, Boyolali, dan Sragen. Kesimpulan: (1) penelitian berhasil mengembangkan tiga paket soal *Ordered Multiple Choice* (OMC) untuk mendeteksi kesulitan belajar peserta didik pada materi Stoikiometri kelas X; (2) butir-butir soal OMC pada paket A, B, dan C memiliki validitas konstruk yang baik dengan *Goodness of Fit* (GoF) berdasar Smart PLS > 0,36 yaitu 0,437; 0,466 dan 0,433. Instrumen memiliki reliabilitas tinggi berturut-turut 0,79; 0,81 dan 0,75; (3) profil peserta didik berupa laporan diagnostik tentang atribut yang sudah dikuasai dan belum dikuasai peserta didik.

Kata kunci: *Attribute Hierarchy Method, Ordered Multiple Choice, Graded Response Model*

DEVELOPING A DIAGNOSTIC INSTRUMENT FOR LEARNING DIFFICULTIES IN CHEMISTRY IN SECONDARY HIGH SCHOOL

¹⁾Sri Yamtinah, ²⁾Budiyono

¹⁾Program Studi Pendidikan Kimia FKIP UNS, ²⁾PPs Universitas Sebelas Maret

¹⁾jengtina_sp@yahoo.com, ²⁾bud@uns.ac.id

Abstract

The purposes of this study are to: (1) develop *Ordered Multiple Choice* (OMC) instruments with a model of *Attribute Hierarchy Method* (AHM) for the diagnosis of stoichiometry learning difficulties in chemistry learning in secondary high school, (2) determine the characteristic of instruments which have been developed based on the *Graded Response Model* (GRM), (3) create a diagnostic profile of learners as an informative report. This development research used the Borg & Gall model. The development of the instrument is done using the AHM with the form of OMC. The determination of attributes and attribute hierarchy was done by the *focus group discussion* (FGD) by three experts, six teachers, and two measurement experts and continued with the Delphi technique by three experts with two rounds. A limited try out was conducted in the high, medium, and low category schools. A feasibility try out was conducted to the high, medium, and low category schools in the regions of Surakarta, Karanganyar, Boyolali, and Sragen. The results are as follows. 1). This research has developed three packages of OMC to detect the learning difficulties of students in the subjects of Chemistry, especially in Stoichiometry of class X. 2). The OMC test items on packages A, B, and C have good construct validity with the *Goodness of Fit* (GoF) greater than 0.36 namely 0.437, 0.466, and 0.433. 3). The learners' profiles are created in the form of diagnostic report about the attributes which have been mastered and have not been mastered by the learners.

Keywords: *Attribute Hierarchy Method, Ordered Multiple Choice, Graded Response Model*

Pendahuluan

Kimia merupakan ilmu yang termasuk ke dalam rumpun IPA, ilmu yang mempelajari mengenai komposisi, struktur, dan sifat zat atau materi dari skala atom hingga molekul serta perubahan atau transformasi serta interaksi mereka untuk membentuk materi yang ditemukan sehari-hari. Pada awalnya kimia merupakan ilmu yang diperoleh dan dikembangkan melalui percobaan (induktif) akan tetapi perkembangan selanjutnya ilmu kimia juga diperoleh dan dikembangkan berdasarkan teori (deduktif). Sebagian aspek kimia bersifat kasat mata (*visible*), artinya dapat dibuat fakta kongkretnya dan sebagian aspek yang lain bersifat abstrak atau tidak kasat mata (*invisible*), artinya tidak dapat dibuat fakta kongkretnya. Namun demikian, aspek kimia yang tidak dapat dibuat fakta kongkretnya harus bersifat kasat logika, artinya kebenarannya dapat dibuktikan dengan logika matematika sehingga rasionalitasnya dapat dirumuskan (Depdiknas, 2003, p.2).

Mata pelajaran kimia di SMA bertujuan mempelajari segala sesuatu tentang komposisi, sifat dan struktur, transformasi, dinamika, dan energetika zat serta penerapannya untuk menyelesaikan masalah sehari-hari. Ilmu kimia merupakan produk (pengetahuan kimia yang berupa fakta, konsep, prinsip, hukum dan teori) temuan saintis dan proses (kerja ilmiah). Oleh sebab itu, dalam proses pembelajaran dan penilaian hasil belajar kimia harus memperhatikan karakteristik ilmu kimia sebagai produk, proses, nilai dan sikap ilmiah, serta aplikasi kimia dalam kehidupan sehari-hari.

Penelitian tentang konsepsi peserta didik dalam kimia didasarkan pada pendekatan konstruktivis untuk belajar, yaitu peserta didik membangun struktur kognitif mereka sendiri. Menurut pendekatan ini, peserta didik menghasilkan makna mereka sendiri berdasarkan latar belakang mereka, sikap, kemampuan, pengalaman sebelum, selama dan setelah proses pembelajaran. Oleh karena peserta didik membangun konsep mereka sendiri, maka bangunan konsepsi mereka seringkali berbeda dengan yang

dibawa oleh guru (Barke, Al Hazari & Yitbarek, 2009, p.2). Konsepsi adalah pemahaman terhadap suatu konsep. Pandangan konstruktivis tentang belajar adalah memperhitungkan ide murid baik dari pengamatan mereka sendiri atau dari bahasa sehari-hari, mengakui bahwa belajar akan melibatkan mengembangkan, memodifikasi dan bahkan menolak ide-ide yang ada, dan menerima pemahaman bahwa pembelajar membangun sesuatu untuk diri mereka sendiri (Bennet, 2007, p.28).

Terdapat dua pendekatan utama dalam konstruktivisme, yaitu konstruktivisme kognitif oleh Piaget dan konstruktivisme sosial oleh Vigotsky. Pada konstruktivisme kognitif, Piaget mengusulkan pembentukan skema atau struktur kognitif, yang merupakan makna, dan pemahaman tentang dunia individu. *Schemata* melayani beberapa fungsi dalam belajar yang mengkategorikan, mengingat, memahami dan memecahkan masalah. *Schemata* atau pengetahuan awal membuat hubungan untuk mengkategorikan pengalaman menjadi lebih efisien untuk diproses. Kategorisasi informasi memfasilitasi proses mengingat (*recall*), dan pemahaman (pemahaman), yang membuat pemecahan masalah lebih produktif. *Schemata* memfasilitasi pemilihan informasi berdasarkan kepentingan kita. Selanjutnya, setelah dipilih, *schemata* memungkinkan material yang dipilih diorganisir secara abstrak dan membantu individu dalam proses menafsirkan dan mengintegrasikan materi baru, berdasarkan apa yang sudah diketahui (Pritchard & Woolard, 2010, p.9). Selama peserta didik membangun atau membangun konsep-konsep mereka sendiri, dimungkinkan sebagian bangunan konsep mereka berbeda dari konsep para ahli.

Burton menyatakan bahwa peserta didik yang diduga mengalami kesulitan belajar, ditunjukkan oleh adanya kegagalan peserta didik dalam mencapai tujuan-tujuan belajar. Peserta didik dikatakan gagal dalam belajar apabila: (1) dalam batas waktu tertentu yang bersangkutan tidak mencapai ukuran tingkat keberhasilan atau tingkat penguasaan materi (*mastery level*) minimal

dalam pelajaran tertentu yang telah ditetapkan oleh guru (*criterion reference*), pada keadaan ini peserta didik disebut *lower group* (2) tidak dapat mengerjakan atau mencapai prestasi sebagaimana semestinya, dilihat berdasarkan ukuran tingkat kemampuan, bakat, atau kecerdasan yang dimilikinya. Peserta didik ini pada kondisi ini dapat digolongkan ke dalam *under achiever*, (3) tidak dapat mewujudkan tugas-tugas perkembangan, termasuk di dalamnya penyesuaian sosial sesuai dengan pola pada fase perkembangan tertentu pada kelompok usianya, kondisi seperti dikatakan siswa sebagai *immature*; (4) tidak berhasil tingkat penguasaan materi yang diperlukan sebagai prasyarat bagi kelanjutan tingkat peserta didik berikutnya, kondisi seperti ini disebut siswa *slow learners* (Burton, 1952, p.135).

Asesmen diagnostik adalah suatu pendekatan yang sangat menguntungkan dalam pengukuran. Penilaian diagnostik memberikan peluang melakukan suatu proses pengujian untuk melayani satu tindakan tambahan selain tujuan-tujuan yang telah ditetapkan, dan dapat digunakan untuk mengintegrasikan proses pembelajaran dan penilaian. Lebih lanjut pengujian diagnostik dapat menemukan bahan pelajaran yang bermakna terhadap setiap peserta didik (Mc Glohen, 2004, p.33).

Gorin dalam Leighton & Gierl (2007: 177) menyatakan bahwa tes diagnostik yang baik adalah yang dapat melakukan estimasi secara keseluruhan atas level kemampuan seseorang. Informasi yang spesifik harus dapat diperoleh dari respon peserta didik yang menunjukkan kelemahan atau inkonsistensi pengetahuan dan menunjukkan potret kemampuan peserta didik.

Model kognitif dalam pengukuran pendidikan mengacu pada gambaran sederhana bagaimana manusia menyelesaikan tugas-tugas standar atau tingkat yang rinci dalam memberikan penjelasan dan memberikan informasi tentang kekuatan dan kelemahan peserta didik (Nichols *et al*, 1995, p.4).

Model kognitif untuk *Cognitive Diagnostic Assessment* (CDA) memiliki setidaknya

empat karakteristik yang menentukan. *Pertama*, model ini berisi keterampilan tertentu pada ukuran butir yang tepat karena keterampilan ini harus memperbesar kinerja proses tes kognitif yang mendasari. *Kedua*, keterampilan harus dapat diukur. *Ketiga*, keterampilan harus relevan dengan pembelajaran dan bermakna bagi kelompok luas pemangku kepentingan pendidikan (*stakeholder*), termasuk peserta didik, orang tua, dan guru. *Keempat*, model kognitif sering mencerminkan hirarki keterampilan yang berjenjang dalam domain karena dependensi berbagi proses kognitif dan fungsi dalam jaringan proses, kompetensi, dan keterampilan lebih besar yang saling terkait. Penilaian yang didasarkan pada model kognitif dapat dikembangkan sehingga butir tes secara langsung dapat mengukur keterampilan kognitif tertentu yang meningkat kompleksitasnya, sehingga memungkinkan item yang menguji kinerja peserta didik dapat dihubungkan dengan informasi mengenai kekuatan dan kelemahan kognitif peserta didik dengan menggunakan model psikometrik dan kognitif diagnostik seperti *Attribute Hierarchy Method* (Leighton, *et al*, 2004, p.217; Gierl, *et al*, 2000, p.37).

Attribute Hierarchy Method (AHM) adalah metode psikometri untuk mengklasifikasikan respon jawaban butir dari peserta ujian menjadi satu set pola atribut terstruktur terkait dengan komponen yang berbeda dari model kognitif kinerja tugas. AHM ini didasarkan pada asumsi bahwa kinerja tes tergantung pada satu set kompetensi hierarkis berjenjang yang disebut atribut kognitif. Atribut adalah proses kognitif atau keterampilan dasar yang dibutuhkan untuk memecahkan tes. Peserta didik yang diuji harus memiliki atribut-atribut untuk menjawab butir tes dengan benar. Hirarki memainkan peran mendasar dalam AHM karena mengarahkan pengembangan tes dan panduan analisis sehingga skor tes psikometri memiliki nilai diagnostik. AHM ini dikembangkan untuk mengatasi dua masalah spesifik yang terkait dengan pengembangan model kognitif dan pengenalan pola statistik (Gierl, 2007, p.331; Gierl, Wang & Zhou, 2008).

Attribute Hierarchy Method (AHM) didasarkan atas asumsi bahwa butir-butir tes dapat dijelaskan melalui suatu set atribut berjenjang yang hirarkis. Asumsi dari ketergantungan atribut adalah sesuai dengan penemuan dari riset kognitif (Vosniadou & Brewer, 1992) di mana skill kognitif tidak dapat beroperasi secara terpisah tetapi merupakan suatu jaringan kompetensi yang saling terkait.

Data yang diperoleh dari studi pendahuluan melalui penyebaran angket pada guru-guru kimia di wilayah Surakarta pada bulan Maret-April 2012 menunjukkan bahwa materi Stoikiometri merupakan materi paling sulit. Hal tersebut didukung dengan data yang diperoleh guru kimia SMA Negeri 4 Surakarta, bahwa sebanyak 64,53% peserta didik kelas X dan 67,16% peserta didik kelas XI IPA tidak tuntas pada ulangan harian pokok bahasan Stoikiometri dengan batas tuntas 72 dan 75. Sementara itu, guru kimia SMA Muhammadiyah 1 Surakarta menyatakan bahwa 62,2% peserta didik kelas X tidak tuntas pada pokok bahasan Stoikiometri dengan batas tuntas 67.

Hasil penyebaran angket juga mendapatkan bahwa 72,2% guru menyatakan bahwa pokok bahasan Stoikiometri merupakan pokok bahasan esensial yang sangat diperlukan untuk memahami dan menguasai pokok bahasan-pokok bahasan lain dalam kimia. Namun demikian, sebanyak 90,7% guru menyatakan belum pernah melakukan upaya khusus untuk mendiagnosis melalui tes diagnostik. Sebanyak 9,3% guru menyatakan melakukan upaya diagnosis melalui tes uraian akan tetapi merasa kesulitan dalam memeriksa hasil pekerjaan peserta didik.

Penyebab kesulitan dalam memahami stoikiometri adalah kompleksitas hitungan yang memerlukan pemahaman tentang konsep mol, penyetaraan reaksi, kemampuan aljabar dan menginterpretasikan bahasa menjadi langkah-langkah prosedural untuk memperoleh jawaban yang benar (Okanlawon, 2010, p.108).

Salah satu alasan kesulitan peserta didik dalam memahami kimia adalah berkaitan dengan berbagai tingkat representasi

yang digunakan dalam pembelajaran kimia untuk menggambarkan dan menjelaskan fenomena kimia. Tiga representasi yang relevan untuk memahami konsep-konsep kimia adalah: (1) representasi sifat makroskopik yang menggambarkan sebagian besar fenomena nyata dan terlihat dalam pengalaman sehari-hari peserta didik ketika mengamati perubahan dalam sifat materi (misalnya perubahan warna, pH larutan mengandung air, dan pembentukan gas dan presipitat dalam reaksi kimia), (2) submikroskopis (atau molekuler) representasi yang memberikan penjelasan pada tingkat partikulat yang dalam hal ini digambarkan sebagai terdiri dari atom, molekul dan ion, dan (3) simbolik (atau ikon) representasi yang melibatkan penggunaan simbol-simbol kimia, rumus dan persamaan, serta gambar struktur molekul, diagram, model dan animasi komputer untuk melambangkan hal (Chandrasegaran, Treagust & Mocerino, 2007, p.297). Dalam mempelajari kimia, kemampuan peserta didik untuk memahami peran setiap tingkat representasi dan mentransfer dari satu tingkat menjadi tingkat lain merupakan aspek penting untuk menghasilkan penjelasan yang dapat dimengerti (Rahayu, 2009, p.669)

Di dalam kelas, guru dapat menggunakan respon terbuka peserta didik untuk mendapatkan informasi tentang pemahaman peserta didik mengenai isi materi pelajaran. Namun, tuntutan mengajar banyak peserta didik (pada sekolah menengah) sering menghambat analisis mendalam tentang tanggapan peserta didik untuk butir *open-ended*. Guru yang telah memiliki sertifikat profesi memiliki kewajiban untuk mengajar di depan kelas 24 jam pelajaran per minggu, apabila alokasi mata pelajaran kimia adalah 3 jam per minggu maka guru tersebut harus mengajar sebanyak 8 kelas. Hal ini tentu akan menyulitkan guru untuk dapat melakukan tes diagnostik dengan butir soal uraian.

Penggunaan butir tes *multiple choice* yang memiliki kelebihan efisien dalam pemeriksaan hasil kerja peserta didik dianggap tidak memberikan informasi diagnostik ten-

tang pemahaman lebih cermat. Dengan demikian, sangat penting untuk dibuat sebuah instrumen yang dapat mengadopsi kelebihan butir tes jenis *open-ended* namun dapat juga memanfaatkan kepraktisan butir tes *multiple choice*.

Briggs, *et all* (2006) mengembangkan *Ordered Multiple Choice* (OMC) sebagai instrumen asesmen untuk fungsi diagnostik. OMC dikembangkan dengan mengkombinasikan efisiensi pilihan ganda tradisional dengan kelebihan kualitatif dari respon pertanyaan *open-ended*. Kunci utama pada pengembangan butir OMC adalah pada *construct map*. Istilah lain dari *construct map* adalah *learning progression*. *Construct map* adalah bentuk pendekatan perkembangan untuk mengukur pemahaman peserta didik. *Construct map* digunakan untuk menyatakan uni dimensionalitas pada masing-masing level. Setiap level merefleksikan tingkatan hirarkis dari pemahaman materi (Briggs, *et all*, 2006: 56).

Asumsi dari ketergantungan atribut adalah sesuai dengan penemuan dari riset kognitif (Vosniadou & Brewer, 1992) yang *skill* kognitif tidak dapat beroperasi secara terpisah tetapi merupakan suatu jaringan kompetensi yang saling terkait. Sampai sekarang, AHM lebih banyak hanya diperkenalkan didasarkan pada model IRT dikotomi. Ada asumsi untuk model dikotomis yang butir tes dapat dijawab dengan benar jika dan hanya jika semua atribut yang terlibat dalam butir tes dapat dijawab dengan benar. Ini berarti bahwa kehilangan satu atribut setara dengan kehilangan semua atribut yang diperlukan. Model Politomus dapat memberikan informasi lebih lanjut untuk skor dari dikotomi. Diagnosis yang akurat membutuhkan deskripsi yang kaya hasil ujian tes (Bolt, *et all*, 2004, p. 57).

Pada penelitian ini AHM dipadukan dengan bentuk tes *Ordered Multiple Choice* (OMC) yang menggunakan skor politomus. Tujuan penelitian ini adalah (1) mengembangkan instrumen *ordered multiple-choice* (OMC) dengan model *Attribute Hierarchy Method* (AHM) untuk diagnosis kesulitan belajar Stoikiometri kelas X, (2) menetapkan

karakteristik atau parameter butir dari instrumen yang telah dikembangkan berdasarkan pada *Graded Response Model* (GRM), dan (3) membuat profil diagnostik peserta didik sebagai laporan yang informatif.

Metode Penelitian

Penelitian ini merupakan jenis penelitian pengembangan yang bertujuan menghasilkan perangkat instrumen yang digunakan untuk mendiagnosis kesulitan belajar stoikiometri pada kelas yang besar. Model pengembangan menggunakan modifikasi model R & D dari Borg & Gall (1983, p.771-796).

Penggunaan model ini dengan pertimbangan bahwa model Borg & Gall khusus didesain untuk meningkatkan hasil pendidikan. Hal tersebut sejalan tujuan utama dari penelitian ini. Model penelitian dan pengembangan adalah proses yang digunakan untuk mengembangkan dan memvalidasi produk dalam dunia pendidikan.

Pengembangan diawali dengan *preliminary study* melalui telaah literatur dan penelitian empiris. Hasil dari telaah literatur menunjukkan bahwa materi stoikiometri merupakan konsep yang sulit sehingga di banyak negara pun, materi stoikiometri juga banyak menimbulkan kesulitan pada peserta didik. Hasil telaah literatur tersebut sejalan dengan hasil temuan empiris yang dilakukan di sekolah-sekolah di wilayah eks karesidenan Surakarta. Dari penyebaran angket diperoleh bahwa yang dianggap sebagai materi (pokok bahasan) tersulit adalah stoikiometri.

Setelah menetapkan pokok bahasan Stoikiometri, selanjutnya melaksanakan *focus group discussion* (FGD) dengan enam orang guru kimia SMA, tiga orang dosen kimia dan pendidikan kimia serta dua orang ahli pengukuran. FGD menghasilkan atribut dan hirarki atribut dilanjutkan dengan teknik Delphi yang melibatkan tiga orang pakar kimia dan pendidikan kimia. Teknik Delphi dilaksanakan dua putaran untuk memperoleh kesepakatan tentang atribut dan hirarki atribut dari Stoikiometri.

Penetapan atribut dan hirarki atribut digunakan untuk menyusun matriks A, matriks R dan matriks Qr sebagai dasar untuk menyusun kisi-kisi butir soal. Kisi-kisi butir soal yang disusun direview oleh enam orang guru yang sejak awal terlibat FGD. Setelah terpilih 20 butir soal maka dikembangkan soal-soal dengan bentuk OMC. Butir-butir soal ini direview dan divalidasi oleh enam orang guru kimia dan tiga orang pakar kimia dan pendidikan kimia. Selanjutnya dilakukan ujicoba kelompok terbatas pada 45 peserta didik, masing-masing 15 orang dari SMA kategori tinggi, sedang dan kurang. Setelah melalui revisi berdasarkan masukan peserta didik, selanjutnya dilakukan ujicoba menengah untuk karakterisasi parameter butir soal. Terakhir dilakukan uji feasibility pada 12 sekolah di wilayah Surakarta, Boyolali, Karanganyar dan Sragen.

Subjek uji coba dalam penelitian ini terbagi menjadi beberapa kategori:

- a. Pada uji coba terbatas yang mengungkap aspek keterbacaan dan kecukupan waktu dengan menggunakan 15 siswa masing-masing dari sekolah dengan kategori baik, sedang dan kurang, yaitu SMA Negeri 4 Surakarta, SMA Negeri 7 Surakarta dan SMA Batik 2 Surakarta.
- b. Pada uji coba skala luas menggunakan 1 sekolah kategori baik, 1 sekolah kategori sedang dan 1 sekolah kategori kurang pada 4 wilayah Surakarta yaitu masing-masing 5 kelas dari SMA Negeri 4 Surakarta, SMA Negeri 7 Surakarta dan SMA Batik 2 Surakarta.
- c. Pada uji kelayakan menggunakan data yang berasal dari uji coba skala luas yaitu masing-masing 1 sekolah kategori baik, 1 sekolah kategori sedang dan 1 sekolah kategori kurang yang berasal dari Kota Surakarta, Kabupaten Karanganyar, Kabupaten Sragen dan Kabupaten Boyolali, yaitu, SMA Negeri 1 Sragen, SMA Negeri Gondang Sragen, SMA Negeri Sukodono Sragen, SMA Negeri 1 Boyolali, SMA Negeri 1 Simo Boyolali, SMA Negeri Nogosari Boyolali, SMA Negeri 2 Karanganyar dan SMA Negeri Gondang Rejo Karanganyar.

Pada studi pendahuluan data yang diperoleh dari angket terbuka dan wawancara pada guru dianalisis secara deskriptif untuk menemukan pokok bahasan yang banyak menimbulkan kesulitan peserta didik pada proses pembelajaran. Data pada uji coba terbatas dianalisis secara deskriptif untuk mencari kelemahan soal dan kemudian merevisinya.

Analisis data pada uji coba skala luas dilakukan untuk mendapatkan informasi validitas konstruk, reliabilitas instrumen, tingkat kesulitan butir, daya pembeda butir, kecocokan model, dan fungsi informasi. Analisis data secara luas menggunakan SmartPLS untuk menguji validitas konstruk dan reliabilitas. Untuk pengujian daya pembeda, tingkat kesukaran dan kecocokan model digunakan Parscale dan untuk menguji efektivitas digunakan Anava.

Hasil Penelitian dan Pembahasan

Langkah awal pada pengembangan instrumen ini adalah menetapkan atribut dan hierarki atribut melalui *focus group discussion* (FGD) yang melibatkan tiga orang dosen, enam orang guru kimia SMA dan dua orang pakar pengukuran. Hasil penetapan atribut dan hierarki atribut melalui FGD tersebut disempurnakan melalui teknik Delphi yang melibatkan tiga orang ahli. Teknik Delphi dilakukan sebanyak dua putaran untuk memperoleh kesepakatan tentang atribut dan hierarki atribut.

Hierarki disusun didasarkan pada *learning progression*. Apakah *learning progression* menjadi dasar dari asesmen diagnostik formal atau non formal pada level sempit (kelas) atau level besar (negara), pilihan tersebut berhubungan dengan model psikometri, kesalahan dalam memilih model akan berakibat fatal (Briggs, & Alonzo, 2009). Pengembangan butir-butir *Cognitive Diagnostic Assessment* (CDA) dipengaruhi oleh empat hal berdasar penelitian Gierl & Cui (2008, p.207), yaitu ukuran butir, dapat terukur, relevansi dengan proses pembelajaran, dan keterampilan yang berjenjang.

Tabel 1. Atribut, Nama Atribut dan Hirarki Atribut Stoikiometri

Atribut	Nama Atribut	Hirarki Atribut
A1	Simbol Atom	
A2	Massa atom relatif (Ar) dan massa molekul relatif / massa rumus relatif (Mr)	
A3	Reaksi kimia dan penyetaraannya	
A4	Hukum-hukum dasar Kimia I: Lavoisier, Dalton dan Proust	
A5	Konsep mol	
A6	Rumus empiris (RE) dan rumus molekul (RM)	
A7	Hukum dasar kimia II: Avogadro	
A8	Pereaksi pembatas	

Untuk menentukan hubungan antara atribut dengan hirarki yang menggunakan AHM, matriks *adjacency* dan *reachability* didefinisikan. Hubungan langsung antara atribut yang ditentukan oleh matriks adjacency biner (A) dari order (k, k), di mana k adalah jumlah atribut, seperti bahwa elemen ijth mewakili ketiadaan (yaitu 0) atau kehadiran (yaitu 1) dari sebuah hubungan langsung antara atribut Ai dengan Aj (Tatsuoka, 2009, p.37).

$$\begin{pmatrix}
 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
 \end{pmatrix}$$

Gambar 1. Matriks A

Setiap baris dan kolom pada matriks A menggambarkan sebuah atribut. Pada baris pertama menggambarkan atribut A1, terlihat bahwa pada kolom 2, 3, dan 4 bertanda angka 1 yang artinya bahwa atribut A1 memiliki hubungan langsung dengan atribut A2, A3, dan A4. Pada baris kedua

yang menggambarkan atribut A2, terlihat bahwa hanya memiliki hubungan langsung dengan atribut A5. Demikian seterusnya sampai dengan baris terakhir yaitu baris ke delapan yang menggambarkan atribut A8, tidak memiliki hubungan langsung dengan atribut manapun. Hubungan langsung (*direct connection*) ini bermakna atribut tersebut menjadi prasyarat bagi atribut yang lain.

$$\begin{pmatrix}
 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\
 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
 \end{pmatrix}$$

Gambar 2. Matriks R

Matriks R adalah gambaran hubungan langsung (*direct connection*) dan hubungan tak langsung (*indirect connection*). Seperti halnya matriks A setiap baris dan kolom pada matriks R juga menggambarkan sebuah atribut. Baris pertama menggambarkan atribut A1 dan baris ke delapan menggambarkan atribut A8. Terlihat pada baris pertama, bahwa semua kolom terisi angka 1, yang artinya

adalah bahwa atribut A1 menjadi memiliki hubungan langsung dan tak langsung dengan dirinya sendiri A1 dan dengan atribut A2, A3, A4, A5, A6, A7 dan A8. Sementara itu, pada baris kedua, pada kolom 2 terisi angka 1 dan kolom 5, 6, 7, dan 8 terisi angka 1 bermakna bahwa atribut 2 memiliki hubungan langsung (*direct connection*) dan tak langsung (*indirect connection*) dengan dirinya sendiri dan atribut no 5, 6, 7, 8.

Item pool direpresentasikan sebagai *incidence matrix* (Q) dengan ordo (k,p) di mana k adalah jumlah atribut dan p adalah jumlah kemungkinan butir. *Item pool* merupakan semua kombinasi dari atribut jika atribut-atribut independen satu sama lain. Dengan demikian, menyebabkan beberapa kemungkinan butir soal tidak sesuai dengan hirarki yang telah disusun sebagai model kognitif, untuk itu perlu dilakukan reduksi terhadap kemungkinan butir soal, sehingga terbentuk *reduced incidence matrix* (Qr) dengan menerapkan batasan hierarki atribut seperti yang didefinisikan oleh matriks R.

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Gambar 3. Matriks Qr

Matriks Qr adalah *blue print* dari butir-butir soal yang hendak dikembangkan dan hanya terbatas untuk menyatakan hubungan antara butir soal dengan atribut-atribut yang telah tersusun.

Melalui matriks Qr tersebut disusun kisi-kisi instrumen berupa indikator-indikator soal. Berdasarkan review guru diperoleh sebanyak 44 indikator soal, namun karena sebagian di antaranya saling tumpang tindih sehingga bias diwakili oleh indikator soal

yang lain, maka dipilih 20 butir indikator untuk ditulis soalnya dalam bentuk OMC.

Salah satu contoh butir soal OMC adalah sebagai berikut.

Gas hidrogen dapat dibuat dengan mereaksikan antara logam Aluminium (Ar=27) dengan larutan asam sulfat. Untuk membuat 6,72 liter gas hidrogen (0 °C, 1 atm), maka di antara pernyataan berikut yang paling tepat adalah....

- Perbandingan mol Al : H₂SO₄ :
Al₂(SO₄)₃ : H₂ = 2 : 3 : 1 : 3
- Persamaan reaksi : 2Al(s) + 3H₂SO₄(aq) → Al₂(SO₄)₃(aq) + 3H₂(g)
- Massa Al yang dibutuhkan pada reaksi tersebut = 5,4 gram
- Jumlah mol H₂ yang dihasilkan pada reaksi tersebut = 0,3 mol

Pada butir OMC tersebut, setiap pilihan adalah jawaban benar yang berasal dari urutan langkah mengerjakan. Langkah pertama untuk dapat menyelesaikan soal tersebut adalah apabila peserta didik dapat menyelesaikan persamaan reaksi (opsi B), selanjutnya adalah mengubah volume gas hidrogen menjadi mol (opsi D). Kemudian dengan membuat perbandingan mol (opsi A) dan akhirnya dapat menghitung massa aluminium (opsi C). Tingkat kesulitan pada langkah-langkah tersebut tidak berurutan, artinya option C yang menjadi level tertinggi pada penyelesaian soal tersebut justru adalah langkah paling mudah, yaitu hanya dengan mengkonversikan mol yang diperoleh pada level 3 (opsi A) ke dalam massa aluminium dengan mengalikan mol dan massa atom relative (Ar Al). Diagnosis dilakukan dengan memperhatikan pilihan jawaban peserta didik. Pilihan jawaban tersebut menunjukkan kemampuan atau level yang dimiliki oleh peserta didik.

Untuk membuktikan validitas isi, maka sebanyak enam orang guru dan tiga orang ahli memberikan penilaian dan masukan terhadap butir-butir soal yang telah disusun. Dengan jumlah rater sebanyak 9 orang maka sebuah item dianggap valid secara isi manakala memiliki V sebesar 0.78 atau lebih (Aiken, 1980, p.134). Seluruh

butir soal dikatakan memiliki validitas isi karena memiliki harga V lebih dari 0,78. Butir-butir soal yang telah dibuktikan validitas isinya kemudian diuji keterbacaan melalui uji coba terbatas oleh 45 peserta didik, yang terdiri dari 15 peserta didik sekolah kategori tinggi, sedang dan rendah.

Berdasarkan uji *multivariate normality distribution* menggunakan Lisrel data penelitian ini tidak berdistribusi normal yaitu tidak memenuhi asumsi *multivariate normality* karena nilai p value skewness dan kurtosis $< 0,05$ maka digunakan SmartPLS untuk menganalisis validitas konstruk instrumen ini. Validitas kontrak dari instrumen dapat ditunjukkan melalui *Discriminant Validity*. Pada penentuan *discriminant validity* dari model pengukuran reflektif, indikator dinilai berdasarkan nilai-nilai *cross loading* dengan konstruksya. Jika korelasi konstruk dengan butir pengukuran lebih besar daripada ukuran konstruk lainnya, maka hal ini menunjukkan bahwa konstruk laten memprediksi ukuran pada blok mereka lebih baik daripada ukuran pada blok lainnya dengan demikian dikatakan memiliki *discriminant validity* yang baik (Ghozali, 2008, p.24).

Harga *Goodness of Fit (GoF) Index* instrumen pada SmartPLS ditentukan berdasarkan formula:

$$\text{GoF} = \sqrt{R^2 \times Com}$$

(Tenenhaus, M., et all, 2005:159-205; Vinzi, et al, 2010, p.58)

Berdasarkan perhitungan diperoleh bahwa nilai-nilai GoF pada masing-masing paket A = 0,437; paket B = 0,466 dan paket C = 0,433. Berdasarkan nilai-nilai GoF tersebut, maka dapat dikatakan bahwa masing-masing paket memiliki harga GoF yang tinggi yaitu $> 0,36$ sehingga memenuhi persyaratan sebagai instrumen yang baik (cocok model).

Di samping pembuktian validitas konstruk, perlu juga dilakukan uji reliabilitas konstruk yang dapat diukur menggunakan 2 (dua) macam cara yaitu *internal consistency* dengan menerapkan formula *composite reliabi-*

lity dari Werts, Linn dan Joreskog (Ghozali, 2008, p.25) dan *Cronbach Alpha*. Jika dibandingkan dengan *composite reliability*, *Cronbach Alpha* cenderung *lower bound estimate reliability*, sedangkan formula *composite reliability* merupakan *closer approximation* dengan asumsi estimasi parameter yang akurat. Untuk itu, penelitian ini menggunakan formula *composite reliability* untuk menentukan reliabilitas konstruk dengan menggunakan SmartPLS.

Pada paket soal A, nilai *composite reliability* untuk masing-masing faktor adalah 0,67; 0,80; 1,00; 0,67; 0,78; 0,79; 0,62; 0,77 dan secara keseluruhan adalah 0,79. Dengan demikian, maka seluruh faktor dikatakan memiliki *composite reliability* yang tinggi. Nilai *composite reliability* pada paket soal B berturut-turut adalah 0,77; 0,73; 1,00; 0,70; 0,82; 0,31; 0,80; 0,74 dan secara keseluruhan adalah 0,81. Pada paket B ini terlihat bahwa pada faktor 6 memiliki *composite reliability* yang rendah ($< 0,6$) hal ini disebabkan pada faktor 6 terdapat sebuah soal yaitu pada nomor 11 memiliki loading faktor sangat rendah (-0,361). Akan tetapi, secara keseluruhan, nilai *composite reliability* termasuk kategori tinggi yaitu 0,81. Paket soal C nilai *composite reliability* setiap faktor berturut-turut adalah 0,79; 0,76; 1,00; 0,32; 0,78; 0,75; 0,69; 0,75 dan secara keseluruhan adalah 0,75. Pada paket C ini terlihat bahwa pada faktor 4 memiliki *composite reliability* yang rendah ($< 0,6$) hal ini disebabkan pada faktor 4 terdapat dua butir soal yaitu pada nomor 12 dan 13 yang setiap soal memiliki loading faktor sangat rendah yaitu -0,006 dan -0,161. Namun secara keseluruhan, nilai *composite reliability* termasuk kategori tinggi yaitu 0,75.

Fit statistics juga berguna dalam mengevaluasi model-model untuk data yang memiliki respons sama jika data-data tersebut dikelompokkan berdasarkan parameternya (Mathilda, 2003, p.704). Untuk melihat *butir fit statistics* digunakan analisis dengan Parscale dengan ketentuan bahwa butir soal dikatakan fit jika memiliki nilai prob $> 0,05$. Berdasarkan analisis *butir fit statistics* dari Parscale tersebut dapat dilihat bahwa pada paket soal A terdapat butir-butir yang tidak

fit yaitu pada nomor 7, 10, 12, 17 dan 18; sedangkan untuk paket soal B terdapat butir-butir soal yang dinyatakan tidak fit yaitu pada nomor 7, 12, 13, dan 18. Adapun pada paket soal C terdapat butir-butir soal tidak fit yaitu pada nomor 7, 10, 12, 17, dan 18. Dengan demikian terdapat nomor-nomor yang pada ketiga paket A, B, dan C dinyatakan tidak fit yaitu pada nomor 7, 12, dan 18.

Adapun tingkat kesukaran pada butir-butir soal dilihat dari parameter b1, b2 dan b3. Parameter b1 menunjukkan tingkat kesulitan antara option level 1 dan level 2, parameter b2 menunjukkan tingkat kesulitan antara option level 2 dan level 3, parameter b3 menunjukkan tingkat kesulitan antara option level 3 dan level 4. Semakin tinggi level, maka tingkat kesukaran akan semakin tinggi. Parameter tingkat kesukaran pada semua paket A, B, dan C memiliki nilai $b1 < b2 < b3$. Dengan demikian, terbukti bahwa pada paket-paket soal semakin tinggi level option pemahaman semakin sulit.

Daya pembeda butir soal dilihat dari parameter a yang pada seluruh butir soal pada paket A, B, dan C memiliki nilai positif, kecuali pada 7 dan 15. Option b dan c pada soal no 7 memiliki level kemampuan yang sama, yaitu pada level 2 sehingga pada butir soal nomor 7, level yang dijelaskan adalah 1, 2, 3.

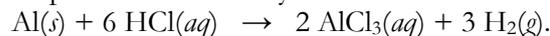
Butir soal nomor 7 yaitu pada faktor 3 dengan atribut persamaan reaksi, terdapat 2 option yang memiliki level sama option b dan c, pada butir soal berikut.

Logam aluminium direaksikan dengan larutan asam klorida encer dan membentuk larutan aluminium klorida dan gas hidrogen. Maka...

- rumus senyawa aluminium klorida adalah $AlCl_3$.
- larutan aluminium klorida dan gas hidrogen sebagai produk.

c. logam aluminium dan larutan asam klorida encer sebagai pereaksi.

d. persamaan reaksinya adalah:



Option b dan c terdapat pada butir soal tersebut. Hal tersebut dimungkinkan menjadi penyebab butir soal nomor 7 memiliki daya pembeda yang rendah yaitu 0,169. Butir soal 15 juga memiliki daya pembeda rendah, hal ini dimungkinkan karena butir soal nomor 15 tersebut juga memiliki *loading factor* yang rendah pada uji validitas konstruk.

Instrumen ini dikemas dalam bentuk *paper and pencil* mengingat kondisi di lapangan yang terbatas kepemilikan sarana komputer dalam jumlah yang memadai. Analisis hasil pekerjaan peserta didik diolah dengan menggunakan bantuan program komputer berbasis Excel. Penggunaan program analisis data dimaksudkan untuk membantu guru lebih cepat dalam menganalisis hasil pekerjaan peserta didik dan menghasilkan output berupa profil diagnostik peserta didik. Program komputer yang dipergunakan untuk analisis hasil pekerjaan peserta didik dibuat dalam bentuk Excel yang sudah cukup familiar di kalangan guru, hal ini dimaksudkan agar guru mudah dalam mengoperasikan.

Uji efektivitas instrumen menggunakan Anova dengan menggunakan SPSS menghasilkan uji beda seluruhnya baik pada paket A, B, dan C berbeda secara signifikan ($P < 0,05$). Artinya instrumen dapat membedakan secara nyata antara peserta didik yang berasal dari sekolah kategori tinggi, sedang dan rendah. Pada pengujian lanjut menggunakan *Scheffe* dan *Bonferroni* untuk paket A, B, dan C terbukti seluruhnya signifikan ($P < 0,05$). Artinya, bahwa kelompok peserta didik kemampuan tinggi > kelompok peserta didik kemampuan sedang > kelompok peserta didik kemampuan rendah.

Nama Siswa: XXX			
No. Soal	SKOR	KEMAMPUAN SUDAH DIKUASAI	KEMAMPUAN BELUM DIKUASAI
1	3	Membentuk dan memberi nama senyawa yang dibentuk dari ion positif bermuatan > 1 dan ion negatif bermuatan 1; membentuk dan memberi nama senyawa yang dibentuk dari ion positif bermuatan 1 dan ion negatif bermuatan >1; membentuk dan memberi nama senyawa yang berasal dari ion-ion dengan muatan positif sama dengan muatan negatif.	Membentuk dan memberi nama senyawa yang berasal dari ion-ion dengan muatan positif dan muatan ion negatif masing-masing > 1 dan berbeda
2	3	Menentukan jumlah atom unsur-unsur yang menyusun sebuah molekul senyawa; memberikan nama suatu molekul senyawa yang diketahui rumus strukturnya; menentukan unsur-unsur penyusun suatu molekul senyawa	Menghitung jumlah atom yang menyusun sejumlah molekul senyawa

Gambar 4. Contoh *Output* Profil Diagnostik

Simpulan dan Saran

Berdasarkan hasil penelitian pengembangan ini dapat dibuat simpulan sebagai berikut.

1. Penelitian ini telah berhasil mengembangkan 3 (tiga) paket soal *Ordered Multiple Choice* (OMC) untuk mendeteksi kesulitan belajar pada pokok bahasan Stoikiometri kelas X. Setiap butir soal OMC terdiri dari pernyataan dan 4 option berisi jawaban benar dengan level pemahaman berjenjang.
2. Butir-butir soal OMC pada paket A, B, dan C yang dikembangkan seluruhnya memiliki indeks validitas isi Aiken lebih dari 0,75 dengan demikian seluruh butir soal memiliki validitas isi yang baik. Paket A, B, dan C memiliki memiliki *Goodness of Fit* (GoF) lebih besar dari 0,36 yaitu 0,437; 0,466 dan 0,433 berdasarkan Smart PLS sehingga memiliki validitas konstruk yang baik. Instrumen memiliki reliabilitas tinggi berdasarkan SmartPLS berturut-turut 0,79; 0,81 dan 0,75.
3. Profil individu peserta didik berfungsi sebagai laporan diagnostik memuat atribut yang dikuasai dan belum dikuasai oleh peserta didik.

Saran

Beberapa hal yang dapat diajukan sebagai saran berdasarkan penelitian ini adalah sebagai berikut.

1. Instrumen *Ordered Multiple Choice* (OMC) yang telah dikembangkan ini lebih tepat dipergunakan bagi peserta didik yang berasal dari sekolah dengan kategori kurang, namun memerlukan petunjuk pengerjaan yang lebih mudah dipahami oleh peserta didik.
2. Sekolah-sekolah yang telah memiliki perangkat komputer yang cukup untuk sejumlah peserta didik, maka akan lebih memudahkan jika perangkat instrumen dikembangkan dalam bentuk *Computer Based Test* (CBT).

Daftar Pustaka

- Aiken, L.R. (1985). Three Coefficient for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 45
- Barke, H.D, Al Hazari & Yitbarek,S (2009). *Misconceptions in Chemistry, Addressing Perceptions in Chemical Education*. German: Springer-Verlag Berlin Heidelberg.

- Bennet, J. (2007). *Teaching and Learning Science*. London: Biddles Ltd., King's Lynn, Norfolk
- Borg, W.R & Gall, M.D.(1983). *Educational Research: An Introduction*, Fourth edition. New York: Longman, Inc
- Briggs, D.C., Alonzo, A.C., Schwab, C., & Wilson, M. (2006). Diagnostik Assessment With Ordered Multiple-Choice Items. *Educational Assessment*, 11(1), 33-63
- Briggs, D.C & Alonzo, A.C (2009). The Psychometric Modeling of Ordered Multiple-Choice Item Responses for Diagnostik Assessment with a Learning Progression. *Paper presented at The Learning Progression in Science (LeaPS) Conference, June 2009, Iowa City*.
- Burton, W.H.(1952). *The Guidance of Learning Activities*. New York: Apleton-Century-Croffts, Inc
- Chandrasegaran, A.L., Treagust, D.F., & Mocerino, M (2007). The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. *Chemistry Education Research and Practice*, 2007, 8 (3), 293-307.
- Depdiknas. 2003. *Kurikulum Berbasis Kompetensi Mata Pelajaran Kimia SMA*. Jakarta : Puskur Depdiknas
- Ghozali, I. (2008). *Structural Equation Modeling: Metode Alternatif dengan Partial Least Square (PLS)*. Badan Penerbit Universitas Diponegoro: Semarang
- Gierl, M.J (2007). Making Diagnostic Inferences About Cognitive Attributes Using the Rule-Space Model and Attribute Hierarchy Method. *Journal of Educational Measurement Winter 2007, Vol. 44, No. 4, pp. 325-340*
- Gierl, M. J., Leighton, J. P., & Hunka, S. (2000). Exploring the logic of Tatsuoka's rule-space model for test development and analysis. *Educational Measurement: Issues and Practice*, 19, 34-44
- Gierl, M.J., Wang, C., & Zhou, J. (2008). Using the Attribute Hierarchy Method to Make Diagnostic Inferences about Examinees' Cognitive Skills in Algebra on the SAT[®]. *Journal of Technology, Learning, and Assessment*, 6(6). <http://www.jtla.org>.
- Gierl, M.J & Cui, Y. (2008). Defining Characteristics of Diagnostic Classification Models and the Problem of Retrofitting in Cognitive Diagnostic Assessment. *Measurement*, 6 : 263-27.
- Gierl, M.J., Cui, Y & Zhou, J (2009). Reliability and Attribute Based Scoring in Cognitive Diagnostic Assessment. *Journal of Educational Measurement Fall 2009, Vol 46, No.3, pp.293-313*
- Hambleton, R.K., Swaminathan, H., & Rogers, H.J (1991). *Fundamental of Item Response Theory*. Sage Publications
- Leighton, J.P & Gierl, M.J (2011). *The Learning Sciences in Educational Assessment: The Role of Cognitive Models*. Cambridge University Press 32 Avenue of the Americas, New York, NY 10013-2473, USA
- Leighton, J.P & Gierl, M.J (2007). *Cognitive Diagnostik Assessment: theory and applications*. Cambridge University Press
- Leighton, J.P, Gierl, M.J & Hunka, S (2004). The Attribute Hierarchy Method for Cognitive Assessment: A Variation on Tatsuoka's Rule-Space Approach. *Journal of Educational Measurement Fall 2004, Vol, 41, No. 3, pp. 205-237*
- Mc Glohen, M.K. (2004). The application of a cognitive diagnosis model via an analysis of a large-scale assessment and a computerized adaptive testing administration. *Disertasi. The university*

- of Texas at Austin. Diakses tanggal 10 Juli 2010, dari scholar.google.com
- Nichols, P.D., Chipman, S.F., & Brennan, R.L. (1995). *Cognitively Diagnostic Assessment*. Lawrence Erlbaum Associate Publishers. Hillsdale, New Jersey
- Okanlawon, A. E. (2010) Constructing a Framework for Teaching Reaction Stoichiometry Using Pedagogical Content Knowledge. *Chemistry*, 19(2), 27-44
- Pritchard, A & Woolard, J (2010). *Psychology for the Classroom: Constructivism and Social Learning*. Routledge 2 Park Square, Milton Park, Abingdon, Oxon.
- Rahayu, S., & Kita, M., (2009). An Analysis Of Indonesian And Japanese Students' Understandings Of Macroscopic And Submicroscopic Levels Of Representing Matter And Its Changes. *International Journal of Science and Mathematics Education (2010) 8*: 667- 688
- Tatsuoka, K.K., (2009). *Cognitive Assessment: An Introduction to the Rule Space Method*. Routledge. Taylor & Francis Group. New York
- Tenenhaus, M., et all (2005). PLS Path Modelling. *Computational Statistics & Data Analysis* 48, 159-205
- Van der Linden, W.J & Hambleton, R.K (1997). Item response theory: Brief history, common models, and extensions. Dalam W.J. Van der Linden & R.K. hambleton (Eds). *Handbook of Modern Item Response Theory* (pp. 1-28). New York: Springer-Verlag Inc
- Vinci, VE., et all. (2010). *Handbook of Partial Least Square: Concept, Methode and Applications*. Springer-Verlag Berlin Heidenberg: Germany
- Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, 24, 535-585.