

## ANALISIS METODE *CHEATING* PADA TES BERSKALA BESAR

Yance Manoppo<sup>1</sup>, Djemari Mardapi<sup>2</sup>

<sup>1</sup>Program Studi Pendidikan Kimia, Fakultas Keguruan dan Ilmu Pendidikan, Unpatti.

<sup>2</sup>Universitas Negeri Yogyakarta

e-mail: ian@chemed-unpatti.info

Diterima 25 Mei 2019/Disetujui 28 Juni 2019

### ABSTRACT

This study aimed to reveal: (1) the characteristics of items of Chemistry Test in National Examination by using the classical test theory and item response theory; (2) the amount of cheating which occurred by using Angoff's B-index Method, Pair 1 Method, Pair 2 Method, Modified Error Similarity Analysis (MESA) Method, and G2 Method; (3) the methods that detect more cheating in the implementation of the Chemistry Test in National Examination for high schools in the year 2011/2012 in Maluku Province. The results of the analysis with the classical test theory approach show that 77.5% items have item difficulty functioning well, 55% items have discrimination yet qualified and 70% items have distractor that works well with the index reliability test of 0,772. The analysis using the item response theory approach shows that 14 (35%) items fit with the model, the maximum function information is 11,4069 at  $\theta = -1,6$ , and the magnitude of the error of measurement is 2,296. The number of pairs who are suspected of cheating is as follows: 13 pairs according to Angoff's B-index Method, 212 pairs according to Pair 1 Method, 444 pairs according to Pair 2 Method, 7 pairs according to MESA Method, and 102 pairs according to G2 Method. The most widely detecting cheating in a row is a Pair 2, Pair 1, G2, Angoff's B-index, and MESA.

**Keywords:** *national examination, items characteristics, methods of cheating.*

### ABSTRAK

Penelitian ini bertujuan untuk mengetahui: (1) karakteristik butir soal Kimia Ujian Nasional berdasarkan teori tes klasik dan teori respon butir; (2) besarnya kecurangan yang terjadi dengan menggunakan Metode *Angoff's B-index*, Metode *Pair1*, Metode *Pair2*, Metode *Modified Error Similarity Analysis (MESA)* dan Metode *G2*; (3) metode yang lebih banyak mendeteksi adanya kecurangan dalam pelaksanaan UN Kimia tingkat SMA/MA Negeri tahun pelajaran 2011/2012 di Provinsi Maluku. Hasil analisis dengan pendekatan teori tes klasik menunjukkan 77,5% butir memiliki tingkat kesulitan butir berfungsi baik, 55% butir daya bedanya belum memenuhi syarat, dan 70% butir memiliki pengecoh berfungsi baik dengan indeks reliabilitas tes 0,772. Analisis dengan pendekatan teori respons butir menunjukkan 14 (35%) butir cocok dengan model, fungsi informasi maksimum 11,4069 pada  $\theta = -1,6$ , dan kesalahan pengukuran 2,296. Jumlah pasangan yang diduga curang adalah: menurut Metode *Angoff's B-index* ada 13 pasangan, menurut Metode *Pair1* ada 212 pasangan, menurut Metode *Pair2* ada 444 pasangan, menurut Metode *MESA* ada 7 pasangan, dan menurut Metode *G2* ada 102 pasangan. Metode yang paling banyak mendeteksi kecurangan secara berturut-turut adalah: *Pair2*, *Pair1*, *G2*, *Angoff's B-index*, dan *MESA*.

**Kata kunci:** *ujian nasional, karakteristik butir, metode kecurangan.*

### PENDAHULUAN

Ujian Nasional (UN) merupakan sistem evaluasi standar pendidikan dasar dan menengah secara nasional. UN dilakukan oleh Depdiknas (Undang-Undang Republik Indonesia nomor 20 tahun

2003). Evaluasi UN dilakukan oleh lembaga yang mandiri secara berkala, menyeluruh, transparan, dan sistematis untuk menilai pencapaian standar nasional pendidikan. Proses pemantauan evaluasi UN harus dilakukan secara berkesinambungan. Pelaksanaan UN diharapkan dapat dipetakan tingkat kemampuan sekolah sehingga dapat menentukan skala prioritas penanganan proses pendidikan. Kejujuran dan prestasi menjadi kata kunci dalam pelaksanaan UN. Untuk mengukur objektivitas, validitas, dan rehabilitas mutu hasil UN, dapat ditelusuri dengan mengembangkan metode analisis pola jawaban peserta ujian. Analisis pola jawaban dapat mengidentifikasi adanya intervensi dari luar atau ketidakjujuran peserta ujian dalam menjawab soal sewaktu penyelenggaraan UN. Salah satu teknik untuk mendeteksi kejujuran dalam pelaksanaan UN yaitu dengan melihat pola jawaban peserta ujian yang bersumber dari Badan Standar Nasional Pendidikan (BSNP).

Kenyataan bahwa sistem pendidikan Indonesia yang menggunakan nilai dari tes atau evaluasi belajar terhadap materi yang diberikan sebelumnya untuk menunjukkan kemajuan dan penguasaan ilmu anak didik, menyebabkan masyarakat memandang prestasi belajar hanya dari pencapaian nilai yang tinggi, bukan pada prosesnya. Pandangan tersebut menimbulkan tekanan pada siswa untuk mencapai nilai yang tinggi. Tekanan yang dirasakan akan membuat siswa lebih berorientasi pada perolehan nilai, bukan pada pemahaman ilmu. Siswa dapat mempersepsi ujian sebagai alat untuk menyusun peringkat dan dapat menyebabkan dirinya mengalami kegagalan, bukan sebagai instrumen yang dapat menunjukkan kemajuan dalam proses belajar.

*Cheating* (ketidakjujuran) pada tes biasanya sering terjadi di semua tingkat dan situasi tes, dari mulai tingkat kelas (Lim & Sean, 2001; Einsenberg, 2004; William, 2002; Anderman, Griesinger & Westerfield, 1998) sampai *high school* atau *academic institution* (Anderman, Cupp & Lane, 2010; Storm & Storm, 2007; Lama, 2008; Chula, Guyette & Piotrowski, 2009; Williams, Craig & Paulhus, 2010; McCabe, Trevino & Butterfield, 2001). *Cheating* berakibat buruk pada rasa keadilan dan kejujuran dari peserta tes serta dapat merusak tujuan tes tersebut di lakukan. Untuk *high-stake assessments* akibat dari ketidakjujuran ini sangat besar karena akan menurunkan tingkat kepercayaan masyarakat terhadap penyelenggaraan tes itu sendiri.

Perilaku *cheating* banyak diakibatkan oleh pengaruh kelompok dimana orang cenderung berani melakukan karena melihat orang lain di kelompoknya juga melakukan. Apabila kecenderungan ini berlangsung secara terus-menerus, maka *cheating* akan menjadi kebiasaan seseorang, yang akan ditransfer tidak hanya pada kegiatan sekolah lainnya tetapi kepada kegiatan kemasyarakatan pada umumnya berdasarkan prinsip *transfer of learning*. Pada akhirnya, *cheating* akan menimbulkan kehancuran di masa mendatang. Hal ini sesuai dengan apa yang dinyatakan Cizek (2001: 3 - 4) sebagai berikut:

*There are many forms of cheating that can occur on such tests. In general, cheating can be defined as any action that violates the rules for administering a test, any behavior that gives an examinee an unfair advantage over other examinees, or any action on the part of an examinee or test administrator that decreases the accuracy of the intended inferences arising from the examinee's test score or performance.*

Berdasarkan pendapat Cizek tersebut, diperoleh bahwa kecurangan secara umum dapat didefinisikan sebagai perilaku yang menyalahi aturan dalam penyelenggaraan tes, perilaku apa pun yang memberikan peserta ujian keuntungan yang tidak wajar atas peserta ujian lainnya, atau tindakan lain dari penyelenggara ujian yang mengurangi keakuratan hasil dimaksud yang berasal dari skor tes atau kinerja peserta ujian.

Penelitian sebelumnya tentang *cheating* sering terfokus pada deskriptif isu-isu seperti gender dan perbedaan budaya. Sebagai contoh, Newstead *et al.* (Anderman, Griesinger & Westerfield, 1998: 84-85) melaporkan bahwa di antara mahasiswa, pria melaporkan kecurangan lebih daripada wanita, siswa yang lebih muda melaporkan kecurangan lebih daripada siswa yang lebih tua, dan siswa kemampuan rendah melaporkan kecurangan lebih daripada kemampuan siswa yang lebih tinggi. Dalam internasional studi tentang kecurangan, Evans, Craig & Mietzel (Anderman, Griesinger &

Westerfield, 1998: 85) menemukan bahwa siswa dari berbagai negara ditafsirkan memiliki kecurangan dalam cara yang berbeda.

Perilaku *cheating* tidak hanya terjadi pada siswa di SMP atau SMA tetapi terjadi pula di bangku kuliah atau universitas. Storm & Storm (2007: 105-106) menyatakan beberapa data yang memprihatinkan adalah survei nasional yang dilakukan oleh *Josephson Institute of ethics* di Amerika pada tahun 2006 dengan responden 36.000 siswa Sekolah Menengah Pertama menemukan 60% siswa menerima dan mengakui pernah melakukan kecurangan pada saat ujian dan pengerjaan tugas. Terjadi peningkatan sebesar 1% dalam kurun waktu 2 tahun, 95% diantaranya mengaku bahwa tidak pernah ketahuan ketika melakukan kecurangan. permasalahan ini dalam berbagai kajian dan penelitian perlu untuk segera mendapatkan penanganan.

Faktor kemajuan teknologi juga memberikan kontribusi terhadap *cheating* pada jalannya UN di sekolah-sekolah. Menurut Deden (2012: 3) yang dimuat pada *Tempo Online* menyebutkan soal UN 2012 untuk Sekolah Menengah Pertama (SMP) di Kota Sukabumi, Jawa Barat, diduga bocor. Hal itu ditandai dengan beredarnya kunci jawaban di kalangan siswa melalui *short message service* (SMS) padahal pihak dinas pendidikan setempat menjamin UN di Kota Sukabumi tidak ada kebocoran. Dengan kecanggihan alat komunikasi, tentunya peristiwa yang sama dapat terjadi di setiap daerah yang telah terjangkau sarana komunikasi seluler ini.

Selain permasalahan bocornya tes, pencapaian kualitas soal sesuai dengan karakteristik butir tes yang baik sangat dipengaruhi sumber daya penyusun tes. Sebagaimana yang diketahui bersama, perangkat tes UN di susun dalam bentuk tes objektif dengan menggunakan *multiple choice* atau bentuk pilihan berganda. Tes seperti ini membutuhkan pemahaman yang komprehensif dan mendalam dari peserta UN sehingga mampu memberikan jawaban yang benar pula. Dengan demikian perangkat tes harus memiliki kehandalan, kesahihan, daya pembeda, distraktor untuk mendeteksi kemampuan siswa secara objektif. Adanya perangkat tes yang kurang baik menyebabkan praktek-praktek *cheating* (kecurangan) lebih mudah dilakukan oleh karena kesulitan siswa dalam menjawab soal dengan benar disamping adanya tekanan untuk harus lulus dalam Ujian Nasional.

Berdasarkan uraian di atas maka yang dimaksud dengan *cheating* dalam tulisan ini adalah segala perbuatan atau trik-trik yang tidak jujur, perilaku tidak terpuji atau perbuatan curang yang dilakukan oleh seseorang untuk mencapai keberhasilan dalam menyelesaikan tugas-tugas akademik terutama yang terkait dengan evaluasi/ujian hasil dengan mengabaikan aturan dan kesepakatan yang ada. Perilaku *cheating* terjadi karena masyarakat memiliki pandangan bahwa prestasi belajar tercermin dari pencapaian nilai yang tinggi, sehingga membuat siswa terpaksa untuk memperoleh nilai tinggi dengan cara apa pun. Masyarakat cenderung semakin permisif sehingga menyebabkan perilaku *cheating* semakin sulit dihilangkan. Fakta-fakta di atas menunjukkan bahwa *cheating* merupakan suatu permasalahan yang menarik untuk dianalisis lebih lanjut. Untuk itu, melalui penelitian ini penulis akan meneliti beberapa metode yang dipakai untuk mendeteksi kemungkinan peserta ujian melakukan *cheating* atau sering diidentikkan dengan *collusion* dengan melihat pola respon jawaban dari peserta tes (*examinees*).

## METODE PENELITIAN

### Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif menggunakan pendekatan *ex-post facto*.

### Waktu dan Tempat Penelitian

Penelitian ini dilakukan di Lab Komputer Pascasarjana Universitas Negeri Yogyakarta setelah memperoleh data respon sampel (LJK) dari Puspendik Kemdikbud Jakarta. Kegiatan penelitian ini dilaksanakan pada Maret 2013.

### Target/Subjek Penelitian

Target dalam penelitian ini adalah perangkat tes UN Kimia tingkat SMA/MA Negeri tahun pelajaran 2011/2012 di Provinsi Maluku.

#### Prosedur

1. Menganalisis butir tes dengan perangkat lunak *IteMan* versi 3.0 untuk melihat karakteristik butir tes yang meliputi tingkat kesukaran butir, daya pembeda butir, dan statistik sebaran jawaban. Selain itu ditentukan juga kehandalan/reliabilitas tes, kesalahan pengukuran (*SE*) dan distribusi skor
2. Menganalisis butir dengan pendekatan teori respons butir menggunakan program *Bilog-MG* 3.0. Dari analisis ini diperoleh parameter butir, parameter peserta, fungsi informasi tes, dan grafik informasi tes. Dari analisis ini diperoleh besarnya Fungsi Informasi maksimum tiap butir, dan skala kemampuan  $\theta$  yang diambil antara -4,0 dan 4,0 dengan interval 0,25.
3. Menaksir kecocokan data dengan model dan parameter butir dengan program *Bilog-MG* 3,0. Kecocokan butir dengan model dapat dilihat dari harga  $p$  yang muncul dari hasil analisis dengan program *Bilog-MG*. Untuk butir yang memiliki harga  $p$  lebih kecil dari 0,01 dianggap tidak cocok dengan model logistik 2-parameter.
4. Mendeteksi *cheating* dengan Metode *Angoff's B-index*, Metode *Pair 1*, Metode *Pair 2*, Metode *MESA*, dan Metode *G2* dengan *software Integrity*<sup>TM</sup>
5. Menentukan keakuratan dari kelima metode deteksi *cheating* dengan melihat metode mana yang dapat mendeteksi *cheating* lebih banyak.

### Data, Instrumen, dan Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah respon peserta UN di Propinsi Maluku khususnya mata pelajaran kimia tingkat SMA/MA sekolah Negeri Paket soal C72 sebanyak 1620 siswa. Pengumpulan data dalam penelitian ini dilakukan dengan menggunakan teknik dokumentasi, yakni mengutip respon siswa (LJK) pada perangkat tes UN Kimia Paket C72 SMA/MA Negeri di Provinsi Maluku tingkat 2011/2012. Data-data berupa LJK diperoleh dari Puspendik Jakarta.

### Teknik Analisis Data

#### 1. Analisis Karakteristik Butir Soal

Analisis terhadap butir soal secara kuantitatif (empiris) menggunakan pendekatan teori tes klasik dan teori respon butir (*IRT*).

##### a. Analisis Empiris dengan Teori Tes Klasik

###### 1) Tingkat Kesukaran Butir Soal

Tingkat kesukaran butir dianalisis dengan menghitung proporsi menjawab benar ( $p$ ) sebuah butir soal. Apabila indeks  $p < 0,30$  maka butir soal termasuk sukar, apabila  $0,30 \leq p \leq 0,70$  maka butir soal memiliki tingkat kesukaran yang sedang, jika  $p > 0,70$  maka soal termasuk soal yang mudah. Akan tetapi oleh karena Ujian Nasional dikembangkan berdasarkan acuan kriteria (tidak membedakan) bukan acuan norma (dapat membedakan), maka besarnya tingkat kesukaran memberikan makna jika minimal 75% (KKM= 75) siswanya bias menjawab butir dengan benar maka siswa tersebut dikatakan sudah "*berhasil*" atau dengan kata lain guru telah berhasil dalam mengajar. Menurut Shapiro (2011: 17) acuan criteria juga dapat digunakan dalam mencari sejauhmana permasalahan yang dihadapi siswa dalam belajar sehingga dapat dicari jalan keluarnya.

###### 2) Daya Pembeda Butir Soal

Daya beda butir soal (*item*) adalah korelasi antara skor butir soal dengan skor total, dihitung dengan rumus korelasi *point biserial*. Besarnya daya beda butir soal untuk menyatakan bahwa butir soal baik adalah minimal 0,3. Sedangkan item soal dengan *point biserial* dibawah 0,3 termasuk item yang kurang baik.

###### 3) Keberfungsian Pengecoh

Untuk mengetahui suatu pengecoh berfungsi secara baik dapat dilihat dari distribusi jawaban. Suatu pengecoh dapat dikatakan berfungsi baik jika paling sedikit pilihan jawaban dipilih oleh 5% peserta tes. Beberapa literatur pada penelitian yang lain juga menyebutkan 2%, tetapi sebenarnya jika ada yang menjawab saja sudah bisa dikategorikan bahwa pengecohnya (*distractor*) sudah berfungsi dengan baik.

4) Indeks Reliabilitas.

Nilai indeks reliabilitas tes dapat dilihat pada nilai koefisien *alpha*. Semakin besar nilai *alpha* menunjukkan semakin *reliabel* tes tersebut dan semakin kecil tingkat kesalahan pengukuran. Standar yang digunakan dalam penelitian ini adalah yang mensyaratkan instrumen yang *reliabel* haruslah memiliki koefisien reliabilitas minimal 0,7.

b. Analisis Empiris Butir dengan Teori Respon Butir

Analisis dengan program *Bilog-MG* menghasilkan *output* dalam bentuk tiga fase. Fase pertama merupakan estimasi butir berdasar teori klasik, fase kedua estimasi parameter butir berdasar Teori Respons Butir dan fase ketiga estimasi kemampuan peserta.

Pada fase pertama diperoleh info tentang banyaknya peserta tes yang menjawab benar, proporsi peluang menjawab benar dibagi peluang menjawab salah serta koefisien korelasi *biserial*. butir yang memiliki nilai koefisien *biserial* negatif dapat mengganggu proses analisis selanjutnya, sehingga butir tersebut tidak diikuti dalam analisis berikutnya. Jumlah butir yang tidak diikuti untuk tahap analisis selanjutnya terdiri dari 2 butir tes Kimia UN SMA/MA Paket C72 tahun pelajaran 2011/2012 Provinsi Maluku.

Fase kedua, estimasi parameter teori respon butir. Pada fase ini diperoleh tentang informasi parameter butir sesuai dengan model teori tespon butir yang digunakan. Selanjutnya untuk model 1-PL didapatkan estimasi tingkat kesukaran, model 2-PL diperoleh estimasi parameter berupa tingkat kesukaran dan daya beda, sedangkan model 3-PL didapatkan informasi tentang tingkat kesukaran, daya beda, dan tebakan semu. Selain parameter butir, pada fase kedua juga dihasilkan statistik kecocokan suatu butir dengan model atau *goodness of fit statistic*. Model yang digunakan untuk estimasi parameter adalah model logistik yang banyak menerima butir yang cocok.

Secara empiris, kualitas butir ditelaah berdasarkan kecocokan data dengan model dan nilai parameter butir. Kecocokan suatu butir dengan model dapat dilihat dari nilai *chi* kuadrat butir dibandingkan dengan harga kritik distribusi *chi* kuadrat sesuai dengan *dk* butir yang bersangkutan pada taraf signifikansi  $\alpha$ . Butir dikatakan tidak cocok model jika nilai  $\chi^2$  butir lebih besar dari harga distribusi  $\chi^2$  pada nilai kritisnya, sebaliknya butir cocok dengan model jika nilai  $\chi^2$  item lebih kecil atau sama dengan nilai distribusi  $\chi^2$ . Atau dikatakan cocok model jika probabilitas  $\chi^2$  lebih besar dari 0,01.

Berdasarkan perbandingan ketiga model parameter logistik untuk uji kecocokan model, maka model 2-PL pada umumnya lebih banyak menghasilkan butir yang *fit* (cocok). Sehingga untuk keseragaman model analisis, estimasi parameter butir untuk seluruh materi uji dianalisis menggunakan model 2-PL.

Fase ketiga didapatkan estimasi parameter kemampuan peserta dan fungsi informasi tes. Pada penelitian ini informasi yang digunakan hanya fungsi informasi tes, sehingga estimasi parameter kemampuan peserta tidak dianalisis. Besarnya fungsi informasi tes dihitung menggunakan program *Excel*.

Berdasarkan kriteria di atas, untuk menentukan kualitas butir yang baik dengan pendekatan teori respon butir didasarkan pada kriteria: (1) butir cocok model, (2) mempunyai daya beda 0 sampai 2, dan (3) tingkat kesulitan -2 hingga +2 (Hambleton & Swaminathan, 1985 : 36).

## 2. Analisis *Cheating*

Selanjutnya untuk mendeteksi adanya *cheating* maka digunakan software *Integrity™*. Software ini adalah sebuah aplikasi *interface online* yang aman dan dirancang untuk menganalisis data ujian pilihan ganda. Program ini mengevaluasi statistik kejujuran suatu tes dan kejujuran akademik siswa yang mengikuti tes. *Integrity* menggunakan *file data* yang dikirimkan oleh pengguna (*client*) untuk mendapatkan informasi statistik uji secara rinci, serta laporan statistik deteksi penyalinan jawaban secara rinci.

## HASIL PENELITIAN

### 1. Karakteristik Butir Tes Klasik dengan Program *Iteman*

#### a. Statistik Butir Soal

Pada tampilan hasil statistik butir soal dapat diketahui hasilnya untuk soal yang dikategorikan sukar atau tidak mencapai keberhasilan 4 butir, kategori kurang berhasil 5 butir, dan yang mencapai keberhasilan 31 butir. Soal yang memiliki daya beda yang baik 22 butir, tidak baik 18 butir sedangkan untuk pengecoh 28 butir yang tergolong baik.

#### b. Statistik Tes/Skala

Berdasarkan statistik hasil analisis program *Microcat Iteman* versi 3.0 terhadap 40 butir soal Kimia UN Paket C72 tingkat SMA/MA tahun pelajaran 2011/2012 yang direspon oleh peserta ujian sebanyak 1620 orang pada Tabel 1 dapat bahwa peserta rata-rata dapat menjawab dengan benar sebesar 31 butir soal ( $mean = 30,981$ ), berarti rata-rata lebih dari setengah (75%) dari jumlah butir soal dapat dijawab dengan benar oleh peserta ujian, skor tertinggi 38 dan terendah 6. Nilai rerata (30,981) yang berdekatan dengan nilai median (32,00). Rata-rata tingkat kesukaran butir soal pada umumnya sedang ( $mean p = 0,775$ ), cenderung banyak pada kisaran indeks 0,30 sampai 0,70. Butir soal cukup mampu membedakan kelompok siswa berkemampuan tinggi dengan kelompok siswa berkemampuan rendah, hal tersebut dinyatakan oleh rerata indeks daya pembeda 0,331.

Indeks reliabilitas soal 0,772 dapat dinyatakan baik, artinya 77,2% perbedaan skor yang terdapat pada peserta ujian adalah kemampuan mereka yang sebenarnya, sedangkan 22,8% perbedaan skor yang ada merupakan kesalahan pengukuran. Menurut Linn (Djemari Mardapi, 1999: 14), kesalahan pengukuran bisa disebabkan oleh variasi acak dalam diri manusia, faktor lingkungan, subjektivitas pengukur, dan alat ukur. Kesalahan baku pengukuran pada perangkat soal hasil analisis secara klasik adalah sebesar 2,209.

**Tabel 1.** Karakteristik Tes Kimia UN Paket 72 Tahun Pelajaran 2011/2012 dengan Program *Iteman*

Karakteristik	Nilai
<i>N of Item</i>	40
<i>N of Examinees</i>	1620
<i>Mean</i>	30,981
<i>Variance</i>	21,441
<i>Std. Dev.</i>	4,630
<i>Skew</i>	-1,776
<i>Kurtosis</i>	4,593
<i>Minimum</i>	6,000
<i>Maximum</i>	38,000
<i>Median</i>	32,000
<i>Alpha</i>	0,772

<i>SEM</i>	2,209
<i>Mean P</i>	0,775
<i>Mean Item-Tot.</i>	0,331
<i>Mean Biserial</i>	0,527

Berdasarkan Tabel 1, rerata tingkat kesulitan dan daya pembeda soal yaitu 0,775 dan 0,331 artinya soal tes ini rata-rata dijawab dengan benar butirnya sebesar 77,5% atau dengan kata lain apabila dengan Kriteria Ketuntasan Minimal (KKM) sebesar 75 maka dapat ditafsirkan bawa tingkat keberhasilan guru dalam melaksanakan proses pembelajaran sudah berhasil, serta soal dapat membedakan siswa yang berkemampuan tinggi dan berkemampuan rendah sebesar 33,1%. Hasil analisis soal Kimia UN Paket C72 menunjukkan bahwa tes tersebut memiliki reliabilitas yang baik yaitu diatas 0,70. Kesalahan pengukuran dapat disebabkan oleh beberapa faktor, seperti lingkungan ketika ujian yang kurang mendukung, kondisi psikologis maupun biologis peserta tes atau terjadi kecurangan ketika pelaksanaan ujian seperti saling kerja sama antara peserta atau adanya bocoran kunci jawaban.

Berdasarkan analisis pogram *Iteman*, soal UN Kimia Paket C72 memiliki kualitas baik, karena 31 (77,5%) butir soal memenuhi kriteria dan 9 (22,5%) butir soal yang tidak memenuhi kriteria. Hal ini didukung juga dengan 70% pengecoh bekerja dengan baik meskipun dari kategori daya beda hanya 55%. Hasil temuan dari 18 butir tersebut terdapat 2 butir soal dengan tanda "*check the key*" dari *print out* program *Iteman*, yaitu butir soal nomor 21 dan 39. Tanda tersebut merupakan suatu peringatan bagi pembuat soal untuk melihat kembali pada pilihan jawaban apakah sudah tepat atau belum. Selanjutnya dari 40 butir soal Kimia Paket soal C72 yang diujikan secara nasional tahun pelajaran 2011/2012 di Provinsi Maluku ditemukan 9 butir soal yang perlu diperiksa kembali sebelum digunakan. Dengan demikian, berdasarkan kriteria yang telah ditetapkan dalam penentuan penerimaan butir soal, maka butir soal yang tidak sesuai dengan teori tes klasik adalah sebanyak 5 (12,5%) butir perlu direvisi dan 4 (10%) butir perlu diganti.

## 2. Karakteristik Butir Soal berdasarkan Pendekatan *IRT*

Pada analisis fase pertama, nilai *biserial* digunakan untuk menentukan apakah suatu butir diikutkan pada analisis berikutnya atau tidak. Fase pertama hasil *output Bilog-MG* menunjukkan bahwa dari 40 butir yang dianalisis terdapat 16 butir soal dengan kategori yang tidak baik atau sukar (butir soal nomor 3, 21, 22, 33, 36, 39) sedangkan sisanya adalah berada pada kategori sedang dan mudah. Berdasarkan daya beda butir tes terdapat 16 butir soal dengan kategori tidak baik (butir soal nomor 3, 9, 11, 14, 18, 19, 21, 22, 26, 27, 28, 33, 36, 38, 39 40), dan 24 butir soal berkategori baik (butir soal nomor 1, 2, 4, 5, 6, 7, 8, 10, 12, 13, 15, 16, 17, 20, 23, 24, 25, 29, 30, 31, 32, 34, 35, 37). Dalam konteks UN yang beracuan kriteria (*criterion referenced test (CRT)*) butir soal dengan tingkat kesulitan mudah dapat digunakan.

Butir soal yang tidak diikutkan dalam fase pertama dapat dilihat pada Tabel 2 berikut.

**Tabel 2.** Butir Hasil Analisis *Bilog* Fase Pertama yang Tidak Dapat Dilanjutkan ke Fase Kedua

No	<i>Correlation</i>	
	<i>Biserial</i>	Keterangan
3	0,062	Tidak Dilanjutkan
9	0,208	Tidak Dilanjutkan
11	0,263	Tidak Dilanjutkan
14	0,268	Tidak Dilanjutkan
18	0,234	Tidak Dilanjutkan
19	0,281	Tidak Dilanjutkan

21	0,042	Tidak Dilanjutkan
22	0,08	Tidak Dilanjutkan
26	0,246	Tidak Dilanjutkan
27	0,271	Tidak Dilanjutkan
28	0,287	Tidak Dilanjutkan
33	-0,032	Tidak Dilanjutkan
36	0,12	Tidak Dilanjutkan
38	0,249	Tidak Dilanjutkan
39	-0,044	Tidak Dilanjutkan
40	0,2	Tidak Dilanjutkan

Pada fase kedua hasil analisis model teori respons-butir dapat dikemukakan bahwa dari 24 butir soal yang dianalisis terdapat 14 butir soal yang baik atau 35 % dan 10 butir soal yang tidak baik atau 25%. Butir soal yang baik adalah butir soal dengan tingkat kesukaran soal ( $b$ ) terletak antara (-2, 2), daya pembeda soal ( $a$ ) terletak antara (0, 2).

Hasil analisis estimasi parameter butir soal UN Kimia paket C72 SMA/MA menggunakan model 2-PL. Hasilnya, jumlah butir soal kimia yang layak dianalisis menggunakan IRT adalah sebanyak 24 butir. Rerata tingkat kesukaran soal kimia ditunjukkan oleh indeks -2,150 yang cenderung mengarah ke soal dengan kategori yang mudah. Rerata daya pembeda ditunjukkan oleh indeks 0,807 yang berarti memiliki daya pembeda yang kurang baik.

Pada Tabel 3 diperoleh informasi bahwa seluruh butir yang dianalisis lanjut ke fase kedua terdapat butir cocok dengan model 2-PL yang berjumlah 19 butir dengan memiliki nilai parameter daya pembeda berada pada interval 0 sampai 2. Informasi ini berarti kurang dari setengah dari keseluruhan butir soal Kimia UN paket C72 memiliki cukup kemampuan untuk membedakan tingkat kemampuan peserta ujian.

**Tabel 3.** Karakteristik Butir Soal UN Kimia Paket C72 dengan IRT

Butir	$a$	$b$	Kecocokan model	Ket
1	0,468	-1,910	Cocok	Baik
2	0,492	-4,105	TC	TB
4	0,727	-2,689	Cocok	TB
5	0,716	-2,474	Cocok	TB
6	0,572	-2,863	Cocok	TB
7	0,587	-3,432	TC	TB
8	1,174	-1,596	Cocok	Baik
10	0,808	-1,990	Cocok	Baik
12	0,805	-1,040	Cocok	Baik
13	0,792	-1,628	Cocok	Baik
15	0,409	-3,551	TC	TB
16	0,763	-1,912	Cocok	Baik
17	0,978	-1,694	Cocok	Baik
20	0,887	-1,947	Cocok	Baik
23	0,764	-1,896	Cocok	Baik
24	1,279	-1,250	TC	TB



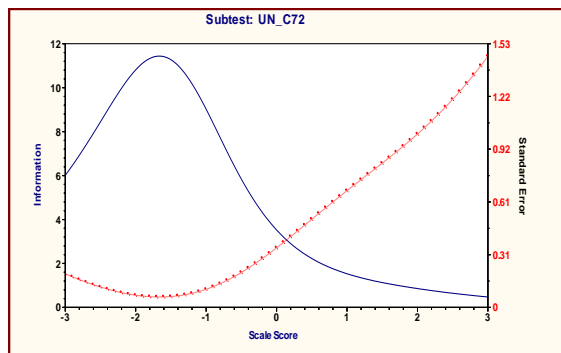
25	0,592	-1,883	Cocok	Baik
29	1,417	-1,749	Cocok	Baik
30	1,035	-1,893	Cocok	Baik
31	1,280	-1,273	Cocok	Baik
32	0,507	-2,387	Cocok	TB
34	1,056	-1,827	Cocok	Baik
35	0,548	-2,146	Cocok	TB
37	0,703	-2,458	TC	TB

Keterangan : TC=Tidak Cocok, TB=Tidak Baik

Hasil estimasi parameter tingkat kesukaran menunjukkan terdapat 15 butir yang berada pada interval -2 sampai 2, yang berarti memiliki tingkat kesukaran yang baik. Sebanyak 9 butir soal memiliki indeks kesukaran dibawah -2,0, sehingga dapat dikatakan sebagai butir yang mudah.

Hasil estimasi parameter di atas dapat dijadikan informasi untuk menentukan karakteristik butir yang baik. Butir yang baik hendaknya memenuhi 3 persyaratan yaitu, butir cocok model, mempunyai daya beda 0 sampai 2, dan tingkat kesulitan -2 sampai +2. Berdasarkan kriteria tersebut maka dapat dikatakan bahwa untuk memilih soal yang baik terdapat 14 butir atau 35% butir soal yang memenuhi persyaratan sebagai butir yang baik. Butir soal tersebut adalah butir nomor 1, 8, 10, 12, 13, 16, 17, 20, 23, 25, 29, 30, 31, dan 40.

Estimasi parameter butir dengan model 2 parameter logistik pada tes Kimia UN paket C72, menghasilkan nilai informasi tes sebesar 11,4069 dengan tingkat kesalahan pengukuran 0,2961. Nilai tersebut tercapai jika parameter kemampuan siswa ( $\theta$ ) sebesar -1,60. Perhitungan fungsi informasi tes dihitung dengan bantuan program *Excel 2007 for windows*. Grafik nilai fungsi informasi tes dengan parameter kemampuan peserta yang sesuai disajikan dalam Gambar 1 berikut.



**Gambar 1.** Grafik Nilai Fungsi Informasi, Skor Kemampuan dan Kesalahan Baku

Kesalahan pengukuran perangkat UN Kimia Paket C72 tingkat SMA/MA Negeri tahun pelajaran 2011/2012 di Provinsi Maluku berdasarkan hasil analisis menunjukkan bahwa besarnya kesalahan pengukuran sebesar 0,296. Kesalahan pengukuran dipergunakan untuk memahami kesalahan yang bersifat acak yang mempengaruhi skor peserta yang sebenarnya dalam pelaksanaan tes. Hasil perhitungan kesalahan pengukuran, diperoleh kesalahan pengukuran sebesar 0,296 dan skor maksimum yang dapat dicapai oleh peserta adalah 38, maka skor peserta UN Kimia Paket C72 tingkat SMA/MA tahun pelajaran 2011/2012 di Provinsi Maluku sebenarnya berkisar pada  $X_i \pm 0,296$ .

### 3. Analisis *Cheating*

Deteksi *cheating* dalam penelitian ini menggunakan lima metode antara lain metode *Angoff's B-index*, metode *Pair 1*, metode *Pair 2*, metode *MESA* dan metode *G2*.

**Tabel 4.** Hasil *Cheating* pada Kelompok Kota (Ambon dan Tual) Berdasarkan Kedekatan Tempat Duduk

Kategori	Metode <i>Cheating</i>				
	<i>B-Index</i>	<i>Pair 1</i>	<i>Pair 2</i>	<i>MESA</i>	<i>G2</i>
Tinggi	0	0	14	0	0
Sedang	0	1	15	0	1
Rendah	0	6	23	0	4
Jumlah	0	7	52	0	5

Hasil keseluruhan *cheating* untuk kelompok kota dapat dilihat pada Tabel 4 yang menyajikan frekuensi *cheating* masing-masing metode pada pelaksanaan UN Kimia SMA/MA untuk paket tes C72 berdasarkan tiga kategori tingkat *cheating* yakni tinggi, sedang dan rendah. Metode *Pair 2* yang memiliki frekuensi terbanyak dalam *cheating* ini setelah metode *Pair 1* dan *G2* sedangkan metode *B-Index* dan *MESA* tidak ada pasangan yang didapati melakukan *cheating*. Besarnya frekuensi kasus *cheating* pada metode *Pair 2* menerangkan bahwa terdapat sebanyak 52 pasangan yang terindikasi melakukan *cheating* dengan model jumlah soal terbanyak yang memiliki jawaban yang sama pada soal-soal berurutan. Sebanyak 14 pasangan memiliki intensitas melakukan kecurangan yang tinggi diikuti 15 pasangan pada level sedang sedangkan sisanya sebanyak 23 pasangan pada level yang rendah. Adanya kemungkinan penyebaran kunci jawaban pada sekolah-sekolah di kota Ambon dibandingkan pada kota Tual juga sangat besar hal ini dapat dilihat pada tingginya frekuensi dari data *benchmark group* pada Tabel 5.

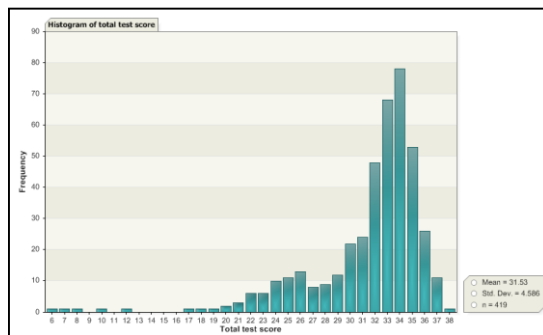
**Tabel 5.** Hasil *Cheating* pada Kelompok Kota Ambon Berdasarkan *Benchmark Group*

Kategori	Metode <i>Cheating</i>				
	<i>B-Index</i>	<i>Pair 1</i>	<i>Pair 2</i>	<i>MESA</i>	<i>G2</i>
Tinggi	0	0	4	0	0
Sedang	0	0	159	0	0
Rendah	0	3	14	0	1
Jumlah	0	3	177	0	1

Data *benchmark group* pada Tabel 5 menjelaskan bahwa sebanyak 177 pasangan pada sekolah-sekolah yang tersebar di kota Ambon memiliki jawaban yang seragam yang mengindikasikan kemungkinan terjadi penyebaran kunci jawaban. Hasil analisis *cheating* pada pelaksanaan UN Kimia SMA/MA Negeri Paket tes C72 di Provinsi Maluku untuk kelompok kota dengan memperhitungkan keseluruhan pasangan yang diduga melakukan *cheating* dengan pasangan terdekat atau yang termasuk dalam *benchmark group* adalah berjumlah 16 pasangan pada metode *Pair 1*, 239 pasangan pada metode *Pair 2*, 6 pasangan untuk metode *G2*, sedangkan

untuk metode *Angoff's B-index* dan metode *MESA* tidak terdapat pasangan yang melakukan *cheating*.

Jika kita memperhatikan Gambar 2 yang memperlihatkan histogram dari total skor tes pada salah satu kelompok kota yakni kota Ambon dapat kita katakan bawa rata-rata (*mean*) memiliki skor 31,53 yang menunjukkan rata-rata peserta tes memiliki kemampuan yang tinggi dengan kemampuan menjawab soal cukup baik.



**Gambar 2.** Grafik Total Sekor Tes Peserta untuk Kota Ambon

Dengan adanya *cheating* yang terjadi pada beberapa pasangan peserta tes membuktikan bahwa tidak semua peserta mampu menjawab dengan kemampuannya sendiri tetapi kemungkinan diduga terjadi kecurangan sewaktu pelaksanaan UN. Hasil analisis *cheating* untuk daerah luar kota yang meliputi 9 Kabupaten yakni Kabupaten Maluku Tengah, Buru, Buru Selatan, Maluku Tenggara Barat, Maluku Barat Daya, Maluku Tenggara, Seram Bagian Barat, Seram Bagian Timur dan Kepulauan Aru. Tabel 6 menyajikan frekuensi *cheating* masing-masing metode pada pelaksanaan UN Kimia SMA/MA untuk Paket tes C72 pada daerah diluar kota dengan frekuensi terbesar pada metode *Pair 2* sebanyak 392 pasangan diikuti dengan metode *Pair 1* sebanyak 205 pasangan, metode *G2* sebanyak 97 pasangan, metode *B-Index* sebanyak 13 dan metode *MESA* sebanyak 7 pasangan.

**Tabel 6.** Hasil *Cheating* pada Kelompok Luar Kota Berdasarkan Kedekatan Tempat Duduk

Kategori	Metode <i>Cheating</i>				
	<i>B-Index</i>	<i>Pair 1</i>	<i>Pair 2</i>	<i>MESA</i>	<i>G2</i>
Tinggi	0	0	245	0	11
Sedang	0	55	53	1	60
Rendah	13	150	94	6	26
<b>Jumlah</b>	<b>13</b>	<b>205</b>	<b>392</b>	<b>7</b>	<b>97</b>

Tabel 6 memberikan informasi bahwa kategori *cheating* tertinggi dalam metode *Pair 2* yang mengindikasikan sebanyak 245 pasangan kemungkinan melakukan *cheating* sangat besar sekali dengan memperhatikan jumlah soal terbanyak yang memiliki jawaban yang sama pada soal-soal berurutan. Sebanyak sedangan metode *Pair 1* yang mengindikasikan tingkat penyalinan jawaban pada pasangan terdekat tergolong sedang dan rendah. Sementara metode *G2* yang mendasarkan pada jumlah jawaban yang benar dan jumlah jawaban yang salah hasil salinan pada

pasangan yang terlibat *cheating*. Hasilnya sebanyak 97 pasangan yang terlibat *cheating* untuk metode ini.

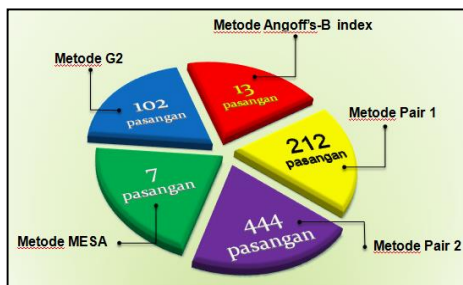
Metode *B-Index* membandingkan jumlah jawaban salah untuk semua pasangan peserta ujian yang masuk dalam interval sama dari hasil tes. interval ini didasarkan pada perkalian jumlah jawaban yang salah untuk sepasang peserta ujian kemudian diamati apakah signifikan berbeda dari nilai rata-rata jawaban salah untuk semua peserta ujian pada interval itu. Frekuensi yang terdeteksi melalui metode ini adalah sebanyak 13 pasangan dengan kategori yang rendah. Selain metode yang sudah disampaikan di atas terdapat metode *MESA* yang mendeteksi sebanyak 7 pasangan yang terlibat *cheating* dengan 1 pasangan pada kategori sedang dan 6 pasangan pada kategori rendah. Hal ini mengindikasikan bahwa probabilitas pasangan yang memiliki jawaban yang sama tetapi salah terdapat pada 7 pasangan tersebut.

**Tabel 7.** Hasil *Cheating* pada Kelompok Luar Kota Berdasarkan *Benchmark Group*

Kategori	Metode <i>Cheating</i>				
	<i>B-Index</i>	<i>Pair 1</i>	<i>Pair 2</i>	<i>MESA</i>	<i>G2</i>
Tinggi	0	0	352	0	0
Sedang	0	76	64	0	93
Rendah	3	199	109	3	20
Jumlah	<b>3</b>	<b>275</b>	<b>525</b>	<b>3</b>	<b>113</b>

Selanjutnya, kemungkinan terjadi penyebaran jawaban pada 9 kabupaten (luar kota), maka Tabel 8 menjelaskan dengan frekuensi yang berbeda-beda yang dilihat berdasarkan data *Benchmark Group*, metode *Pair 2*, *Pair 1*, *G2*, *B-Index* dan *MESA* menunjukkan menunjukkan frekuensi indikasi penyebaran jawaban sangat besar. Hasil analisis *cheating* pada pelaksanaan UN Kimia SMA/MA Negeri Paket C72 tahun pelajaran 2011/2012 di Provinsi Maluku untuk kelompok luar kota (kabupaten) dengan memperhitungkan keseluruhan pasangan yang diduga melakukan *cheating* dengan pasangan terdekat atau yang termasuk dalam *benchmark group* adalah terdapat 16 pasangan untuk metode *Angoff's B-index*, 480 pasangan untuk metode *Pair 1*, 917 pasangan untuk metode *Pair 2*, 10 pasangan untuk metode *MESA*, dan 210 pasangan untuk metode *G2*.

Secara keseluruhan untuk kelompok kota dan luar kota berdasarkan analisis metode *cheating* ditampilkan pada Gambar 3. Dari grafik tersebut ditampilkan terdapat jumlah pasangan yang diduga melakukan *cheating* untuk metode *Angoff's B-index* terdapat 13 pasangan, metode *Pair 1* terdapat 212 pasangan, metode *Pair 2* terdapat 444 pasangan, metode *MESA* terdapat 7 pasangan, dan metode *G2* terdapat 102 pasangan. Pasangan yang diambil adalah pasangan yang duduknya paling berdekatan sehingga adanya indikasi terjadinya *cheating*. Selanjutnya bagi pasangan yang terindikasi *cheating* tetapi pada lokasi tempat duduk yang sudah berjauhan, pasangan ini dimasukkan kedalam kelompok *benchmark group* untuk menduga terjadinya kebocoran atau penyebaran kunci jawaban pada sekolah maupun pada daerah tersebut.



**Gambar 3.** Grafik Pasangan yang Terindikasi *Cheating* pada UN SMA/MA Kimia Paket C72 Tahun Pelajaran 2011/2012 di Provinsi Maluku

Ketidakjujuran siswa menghadapi UN memupuk ketidakjujuran siswa menghadapi kehidupan bermasyarakat, yang pada akhirnya akan melahirkan generasi-generasi muda yang tidak jujur. Ironisnya adanya kemungkinan terjadinya *cheating* pada UN 2011/2012 ini yang kemungkinan terjadi penyontekan jawaban dan penyebaran jawaban oleh pihak-pihak sekolah (bahkan pemerintah daerah) mengizinkan (bahkan membantu) *cheating* tersebut berlangsung sebagai tujuan agar kelulusannya mencapai 100%.

Akar permasalahan dari *cheating* UN menurut peneliti adalah UN sebagai alat untuk menentukan kelulusan yang hanya di administrasikan satu kali dalam setahun. Bagi pihak-pihak tertentu (siswa, guru, kepala sekolah, orang tua, atau kepala daerah) yang merasa tidak siap menghadapi kemungkinan siswa-siswinya tidak lulus akan tergoda untuk melakukan kecurangan.

## KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan pada perangkat Ujian Nasional Kimia Paket C72 tingkat SMA/MA Negeri tahun pelajaran 2011/2012 di Provinsi Maluku dapat disimpulkan sebagai berikut.

1. Berdasarkan pendekatan teori tes klasik terdapat 77,5% butir memiliki tingkat kesulitan butir berfungsi baik, 55% butir daya bedanya belum memenuhi syarat dan 70% butir pengecoh (*distractor*) berfungsi baik dengan indeks reliabilitas tes 0,772.
2. Berdasarkan pendekatan teori respons butir terdapat 14 (35%) butir soal cocok (*fit*) dengan model, fungsi informasi maksimum 11,4069 pada  $\theta = -1,6$ , dan besarnya kesalahan pengukuran (*SEM*) 2,296.
3. Berdasarkan hasil analisis metode *cheating* terdapat jumlah pasangan yang diduga melakukan *cheating*: metode *Angoff's B-index* terdapat 13 pasangan, metode *Pair 1* terdapat 212 pasangan, metode *Pair 2* terdapat 444 pasangan, metode *MESA* terdapat 7 pasangan, dan metode *G2* terdapat 102 pasangan yang terdeteksi *cheating*.
4. Berdasarkan hasil analisis, metode yang lebih banyak mendeteksi adanya *cheating* dalam Ujian Nasional Kimia secara berturut-turut adalah metode *Pair 2*, metode *Pair 1*, metode *G2*, metode *Angoff's B-index* dan metode *MESA*.

**DAFTAR PUSTAKA**

- Anderman, E. M., Griesinger, T., & Westerfield, G. (1998). Motivation and cheating during early adolescence. *Journal of Educational Psychology*, 90, 84-93.
- Anderman, E. M., Pamela, K. C., & Derek, L. (2010). Impulsivity and academic cheating. *Journal of Educational Psychology*, 90, 84-93.
- Cizek, G. J. (1999). *Cheating on tests: How to do it, detect it, and prevent it*. Mahwah, New Jersey: Lawrence Erlbaum.
- \_\_\_\_\_. (April 2001). An overview of issues concerning cheating on large-scale tests. *Paper presented at the annual meeting of the National Council on Measurement in Education*, in Seattle, WA
- Chula, G. K., Roger, W.G., & Chris, P. (2009). Online exams and cheating: an empirical analysis of business students' views. *The Journal of Educators Online*, 6, 1.
- Deden, A. A. (September 2012). Soal Ujian Nasional SMP Diduga Bocor. *Tempo Online* Diambil pada tanggal 30 September 2012, dari : <http://www.tempo.co/read/news>
- Depdiknas (2003). Undang-Undang RI Nomor 20, Tahun 2003, tentang Sistem Pendidikan Nasional.
- \_\_\_\_\_. (2005). Peraturan Pemerintah RI Nomor 19, Tahun 2005, tentang Standar Nasional Pendidikan.
- Djemari Mardapi. (1999). *Estimasi kesalahan pengukuran dalam bidang pendidikan dan implikasinya pada ujian nasional*. Yogyakarta: UNY.
- Eisenberg, J. (2004). To cheat or not cheat: effects of moral perspective and situational variables on students' attitudes. *Journal of Moral Education*, 33, 2, 163-178.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: principles and applications*. Boston, MA: Kluwer Academic Publishers.
- Lama, M. Al. (2008). Student's attitudes toward cheat and relation to demographyc factors. *European Journal of Social Science*, 7, 1, 140-146.
- Lim, V. K. G., Sean, K. B. S. (2001). Attitudes toward, and intentions to report, academic cheating among students in Singapore. *Ethnic & Behavior*, Lawrence Erlbaum Associates, 11(3), 261-274
- McCabe, D. L., Linda K. T., & Kenneth, D.B. (2001). Cheating in academic institutions: a decade of research. *Ethnic & Behavior*, 11(3), 219-232.
- Shapiro, E. S. (2011). *Academic skill problems* (4th Ed.). New York: The Guildford Press.
- Williams, J. B. (2002). The plagiarism problem: are students entirely to blame. *Proceedings of ASCILITE*. Australia
- Williams, K., M., Craig N., Delroy L.P. (2010). Identifying and profiling scholastic cheaters: their personality, cognitive ability, and motivation. *Journal of Experimental Psychology*, 16 (3), 293-307.