

**ANALISIS KUANTITATIF TES BUATAN GURU GEOGRAFI  
BERDASARKAN MODEL LOGISTIK DUA PARAMETER (L2P) TEORI  
RESPONSI BUTIR DI KABUPATEN PADANG PARIAMAN**

**ZAMHAR BAKRI**

**Dosen Pendidikan Geografi STKIP Ahlussunah**

[zamharbakri@gmail.com](mailto:zamharbakri@gmail.com)

**Article History**

**Received : March 2019**

**Accepted : June 2019**

**Publish : July 2019**

**Keyword :**

*Test Made In Teachers  
And Logistic Models Of  
Two Parameters (L2p)  
Theory Of Grand  
Response*

**ABSTRACT**

*The type of research used was descriptive quantitative with the population in the form of items tested to participants who took the exam from January to June 2013/2014, while the sample as respondents was class X SMA Pariaman Regency, amounting to 1012 people, Data collection techniques namely the study of documents, in the form of LJK data for participants of the January-June 2013/2014 exam along with the answer key. Before the item is carried out, it is necessary to test the TRB requirements, namely in the form of dimensions with factor analysis. Parameter invariance problem with correlation analysis and local independence with covariance analysis. After fulfilling the three requirements, then it can be determined the level is difficult by using the Ascal 3.2 program and the function of item information.*

*The test results on the requirements of the TRB obtained that (1) the results of the validity test there are 33 valid questions and 17 items that drop, (2) the reliability of the test with the results are not reliable with a value of 0.55 (3) the local independence is the smallest covariance index and close to zero in the diagonal cell of the matrix. Analysis of the level of difficulty and power difference found that there were 33 items (66%) acceptable, 17 items (34%) were rejected. The value of the information function is good, because it is difficult for most of the levels to be at  $(-2 \leq b_i \leq 2)$ . Based on the findings of this author, it is recommended that the choice of answers, pictures and sentences used be corrected, and in accordance with SKL.*

## ABSTRAK

Jenis penelitian yang digunakan adalah deskriptif kuantitatif dengan populasi berupa butir-butir soal yang diujikan kepada peserta yang mengikuti ujian semester Januari-Juni 2013/2014, sedangkan yang menjadi sampel sebagai respondenya adalah kelas X SMA Kabupaten Padang Pariaman yang berjumlah 1012 orang, Teknik pengumpulan data yaitu studi dokumen, berupa data LJK peserta ujian semester Januari-Juni 2013/2014 beserta kunci jawaban. Sebelum terhadap butir soal dilakukan, maka perlu pengujian persyaratan TRB yaitu berupa unidimensi dengan analisis faktor. Soal Invariansi parameter dengan analisis korelasi dan independensi lokal dengan analisis kovarian. Setelah terpenuhinya ketiga persyaratan tersebut barulah dapat ditentukan taraf sukar dengan menggunakan program ascal 3.2 dan fungsi informasi butir soal.

Hasil pengujian terhadap persyaratan TRB tersebut diperoleh bahwa (1) hasil tes validitas terdapat 33 butir soal yang valid dan 17 butir yang drop, (2) reliabilitas tes dengan hasil tidak reliabel dengan nilai 0,55 (3) Independensi lokal terdapat indeks kovariansi yang terkecil dan mendekati nol dalam sel diagonal matrik. Analisis taraf sukar dan daya beda diperoleh bahwa terdapat 33 butir soal (66%) dapat diterima, 17 butir soal (34%) ditolak. Nilai fungsi informasi termasuk baik, karena taraf sukar sebagian besar berada pada  $(-2 \leq b_i \leq 2)$ . Berdasarkan hasil temuan penulis ini, disarankan agar pilihan jawaban, gambar dan kalimat yang digunakan diperbaiki, serta sesuai dengan SKL.

## A.PENDAH ULUAN

Pendidikan merupakan hak asasi setiap warga negara Indonesia dan untuk itu setiap warga negara Indonesia berhak memperoleh pendidikan yang bermutu sesuai dengan minat dan bakat yang dimilikinya tanpa memandang status sosial, status ekonomi, suku, etnis, agama, dan gender. Pemerataan akses dan peningkatan mutu pendidikan akan membuat warga negara Indonesia memiliki kecakapan hidup (*life skills*) sehingga mendorong tegaknya pembangunan manusia seutuhnya serta masyarakat madani dan modern yang dijiwai nilai-nilai Pancasila, sebagaimana yang telah diamanatkan dalam UU No 20 Tahun 2003 tentang Sistem Pendidikan Nasional. Serta pengelolaan dan penyelenggaraan pendidikan bertujuan membangun landasan bagi berkembangnya potensi peserta didik agar menjadi manusia yang beriman dan bertakwa kepada Tuhan Yang Maha Esa, berakhlak mulia, berkepribadian luhur, berilmu, cakap, kritis, kreatif, inovatif, sehat, mandiri, percaya diri, toleran, peka sosial, demokratis, dan bertanggung jawab terdapat dalam Peraturan Pemerintah Nomor 17 Tahun 2010.

Kemajuan ilmu pengetahuan dan teknologi saat ini menyebabkan tantangan yang dihadapi oleh sistem pendidikan di Indonesia semakin berat. Tantangan ini

menuntut kerja keras berbagai komponen yang terlibat dalam bidang pendidikan seperti: pemerintah dalam hal ini Kementrian Pendidikan Nasional dan Kebudayaan (Kemendikbud) beserta sekolah dengan unsur pimpinan dan gurunya.

Keluaran pendidikan Indonesia mendatang harus siap bersaing dalam memasuki dunia kerja yang tidak saja di dalam negeri tapi juga di negara-negara lain di dunia. Salah satu upaya pemerintah dalam meningkatkan mutu pendidikan di Indonesia adalah melakukan perubahan kurikulum pada tahun 2004 yang dikenal dengan Kurikulum Berbasis Kompetensi (KBK) ke kurikulum 2006 yang lebih dikenal dengan Kurikulum Tingkat Satuan Pendidikan (KTSP). Saat ini kementrian pendidikan dan kebudayaan sudah menggunakan kurikulum 2013 yang berkarakter. Semua komponen yang berkaitan dengan penggunaan Kurikulum 2013 diharapkan siap untuk menerapkan atau menggunakan kurikulum ini. Kurikulum 2013 merupakan salah satu upaya pemerintah untuk memperbaiki kondisi keterpurukan dunia pendidikan Indonesia saat ini.

Kurikulum 2013 bertujuan untuk meningkatkan mutu pendidikan melalui kemandirian dan inisiatif sekolah dalam mengembangkan kurikulum, mengelola

dan memberdayakan kepedulian warga sekolah dan masyarakat dalam pengembangan kurikulum melalui pengambilan keputusan bersama dan meningkatkan kompetensi yang sehat, satuan pendidikan tentang kualitas pendidikan yang akan dicapai.

Pentingnya sistem evaluasi yang bermutu dikemukakan pula oleh Roses bahwa, dalam rangka memasuki kemajuan dunia dibidang IPTEK. sistem pendidikan mesti menemukan suatu strategi baru yang ia sebut dengan “percepatan belajar di abad ke 21” (*accelerated learning for the 21<sup>st</sup> century*). Satu di antara 16 strategi percepatan pendidikan dan pengajaran itu ialah memperbaiki sistem evaluasi di sekolah (*modify the evaluation system*). Disisi lain Soedijarto (2000:18) mengemukakan pula bahwa, salah satu titik perhatian penting yang perlu dilakukan dalam pengembangan sistem Pendidikan Nasional di Indonesia adalah meningkatkan kemampuan guru dalam bidang evaluasi.

Kebijakan dan beberapa pemikiran di atas, sesuai pula dengan program yang dirintis oleh *The Holmes Goup* (sebuah konsorsium yang beranggotakan hampir 100 orang peneliti Perguruan Tinggi di Amerika) yang tertarik dengan masalah-masalah pendidikan. Mereka menyepakati bahwa dalam memajukan pendidikan, perhatian bukan ditujukan untuk

menciptakan sekolah apa yang diharapkan tetapi bagaimana guru-guru dipersiapkan. Dari pemikiran ini dapat dipahami bahwa, pengembangan sekolah pada dasarnya adalah pengembangan kualitas para gurunya. Salah satu upaya untuk mengembangkan kualitas guru menurut konsorsium itu adalah meningkatkan kemampuan serta kinerja guru dalam melaksanakan sistem penilaian dan evaluasi.

Dalin (1998:156) dalam gagasannya tentang beberapa indikator penting bagi pengembangan suatu sekolah adalah dengan meningkatkan guru (*teacher development*) terutama memberi guru kemampuan praktis (*to give the individual teacher practical skills*). Salah satu kemampuan praktis yang dimaksud adalah kinerja guru dalam membuat soal. Kinerja guru dalam mengevaluasi hasil belajar dimulai dari merencanakan materi pembelajaran, membuat Tujuan Pembelajaran Khusus (TPK), membuat kisi-kisi soal, membuat soal itu sendiri, menganalisis butir soal, melaksanakan tes, mengukur, memberi nilai, sampai memutuskan apakah peserta didik sudah berhasil atau belum. Masing-masing kegiatan saling berkaitan satu sama lainnya. Artinya, evaluasi tidak bisa berlangsung tanpa merancang seperangkat alat tes, memberikan tes, mengukur hasil belajar

peserta didik dan memberikan penilaian terhadap hasil belajar.

Hasil wawancara pada tanggal 23 Maret 2014 dengan tiga orang guru mata pelajaran geografi di Sumatera Barat, khususnya pada beberapa tim MGMP Geografi Kabupaten Padang Pariaman, diperoleh gambaran bahwa guru cenderung: (1) belum mampu membuat tes atau mengevaluasi peserta didiknya dengan baik, (2) membuat tes ketika diperlukan dan tanpa memikirkan syarat-syarat atau kriteria tes yang benar, (3) tidak begitu memikirkan apakah tes yang dibuat dapat dijawab oleh peserta didik, terlalu mudah atau terlalu sulit.

Hasil wawancara di atas diperkuat oleh jawaban yang diberikan oleh beberapa orang guru bidang studi geografi yang dijangkau melalui sejumlah pertanyaan dalam angket terbuka. Berdasarkan jawaban yang diberikan diperoleh gambaran bahwa: (1) guru hanya mengenal istilah-istilah evaluasi, tetapi belum mengerti tentang pengertian, maksud dan cara melakukannya dengan baik, (2) kurang memahami konsep validitas, reliabilitas, tingkat kesukaran soal, daya beda dan fungsi pengecoh, (3) jarang menyusun soal merujuk kepada pedoman penyusunan soal yang ada, (4) soal cenderung disusun berdasarkan pengalaman dan pengetahuan apa adanya, (5) soal cenderung disusun sesaat

menjelang ujian akan dilaksanakan, (6) hanya memanfaatkan perangkat soal yang dibuat oleh beberapa penerbit buku dan penerbit Lembaran Kerja Peserta didik (LKS) yang belum diuji keterandalannya.

Fakta di atas sesuai dengan pendapat Umar dalam Little dan Wolf (1996) bahwa salah satu masalah dalam sistem penilaian di Indonesia adalah rendahnya kualitas butir soal yang dibuat oleh guru (*The quality of items used is generally poor*). Dijelaskan pula bahwa salah satu cara mengatasi permasalahan ini adalah dengan menyediakan penulis-penulis soal dan penilai yang terlatih dengan baik (*well-trained item writers and reviewers*).

Ada dua hal yang harus diperhatikan dalam pembuatan tes. Pertama, tes tersebut harus dapat mengungkapkan secara benar ciri terpendam yang ada pada peserta tes, Kedua, tes harus dapat mengungkapkan tanggapan dari peserta tes dengan informasi yang berhubungan (Naga, 1992:6).

Pada penelitian ini akan menggunakan teknik analisis tes modern yaitu Teori response Butir (*Item response theory*) disingkat dengan TBR. Teori ini sudah lama berkembang, akan tetapi kurang diminati, padahal TBR mempunyai banyak kelebihan di antaranya dapat dilakukan dengan komputerisasi, dimana semua dapat dilakukan dengan mudah dan serba cepat. sementara pada teori klasik terdapat

beberapa keterbatasan, ciri teori klasik ditunjukkan oleh kenyataan bahwa parameter butir tes tidak dapat dipisahkan dari parameter peserta tes. Artinya kelompok parameter butir yang sama diisi oleh kelompok parameter peserta yang sama. Jika kelompok parameter butir yang sama diisi oleh kelompok parameter peserta yang berbeda, maka ciri atau karakteristik butir itu berubah. Dengan kata lain taraf sukar butir dan daya beda butir berubah bila dijawab oleh kelompok peserta yang berbeda.

Demikian juga sebaliknya, bila kelompok parameter yang sama menjawab butir tes yang berbeda, maka ciri atau karakteristik kelompok parameter peserta berubah, artinya kemampuan mereka berubah karena mengisi butir tes yang berbeda. Pada teori modern bertujuan menghilangkan kelemahan pada teori klasik. Teori modern memisahkan kedua parameter ini, yaitu parameter butir tes dengan parameter peserta tes. Dengan adanya teori modern ciri parameter butir akan tetap sama, walaupun dijawab oleh peserta yang berbeda, demikian pula sebaliknya ciri parameter peserta akan tetap sama walaupun diuji dengan butir apapun. TBR memiliki 3 model Logistik, yaitu model Logistik Satu Parameter (L1P), logistik Dua Parameter (L2P), dan logistic

Tiga Parameter (L3P) tergantung pada berapa banyak parameter yang digunakan.

Penelitian ini menggunakan model Logistik dua Parameter (L2P) yaitu untuk melihat berapa taraf sukar dan fungsi informasi t Ujian Semester Januari-Juni 2013/2014. Dengan demikian penelitian mengangkat penelitian ini dengan judul “Analisis Kuantitatif Tes Buatan Guru Geografi berdasarkan Model Logistik Dua Parameter (L2P) Teori Responsi Butir di Kabupaten Padang Pariaman”

Hill (1994:20) mengemukakan bahwa, evaluasi adalah proses pengambilan kesimpulan memberi interpretasi dan membuat keputusan yang profesional berdasarkan informasi yang terkumpul. Hill dan Ruptic juga menjelaskan evaluasi sebagai proses yang mengacu pada suatu aktivitas, pengumpulan data yang sistematis, analisis dan pelaporan informasi yang dapat digunakan untuk merubah perilaku untuk meningkatkan suatu proyek atau program. Di bidang pendidikan, menurut Nitko (1996:176) Evaluasi adalah proses menetapkan keputusan tentang nilai, hasil atau kinerja siswa (*judgment about student's product or performance*). Sejalan dengan pendapat Nitko, menurut Gronlund (1990) evaluasi adalah suatu proses yang sistematis untuk menentukan apakah peserta didik mencapai tujuan pengajarannya.

Evaluasi yang baik adalah evaluasi yang memenuhi syarat konteks (materi, rencana tujuan, dsb) maupun konstruk (validitas, reliabilitas, tingkat kesukaran soal dan daya beda soal yang baik). Namun demikian menurut Suryabrata (2002:137), sangat sulit mendapatkan suatu butir tes yang memenuhi semua persyaratan. Keputusan tentang suatu butir tes apakah dikategorikan baik atau dipakai tergantung kepada kebutuhan dan kebijakan syarat mana yang diharapkan terpen

10  
Sebelum evaluasi dilakukan, evaluator perlu melakukan beberapa kegiatan yang mendahuluinya yaitu, (1) merancang tes, (2) melakukan penilaian (*assessment*) dan (3) melakukan pengukuran (*measurement*). Nitko (1996:4) mengemukakan evaluator perlu membedakan konsep dalam evaluasi pendidikan.

Kinerja (*performance*) menjadi *trend* dan diterima oleh banyak kalangan termasuk dalam dunia pendidikan. Kinerja menurut Nilson (1997:70) berarti melakukan sesuatu (*performance is doing*). Nicholes dalam Stolovitch dan Erica mengemukakan, pada dasarnya kinerja adalah hasil dari sikap. Meningkatkan kinerja seseorang dapat menghasilkan suatu produksi dan kepuasan kerja mereka. Jadi, kata kinerja itu identik dengan orang yang

tidak bisa diam (*disquiet persons*) dalam menghadapi suatu pekerjaan.

Pendapat yang sama juga dikemukakan oleh Rowland dalam Latief yang mengemukakan kinerja sebagai tindakan seseorang dalam melakukan sesuatu. Kinerja menurut Rowland dapat dilihat dari upaya-upaya yang dilakukan seseorang dalam mencapai target yang ditetapkan. Biasanya kinerja yang rendah ditandai oleh upaya yang sedikit dan sebaliknya kinerja tinggi ditandai oleh upaya yang maksimal. Kriteria kinerja setiap orang didasarkan pada tugas dan tanggung jawab keseharian yang dibebankan atau ditargetkan kepadanya.

Untuk mendapatkan soal yang bermutu memerlukan beberapa kegiatan evaluasi antara lain kegiatan: membuat soal yang valid, membuat soal yang reliabel, membuat soal yang mempunyai taraf sukar yang baik, membuat soal yang mempunyai daya pembeda dan membuat soal yang memfungsikan opsi pengecoh. Tugas ini merupakan tugas rutin yang dilakukan guru untuk mengambil keputusan tentang keberhasilan siswa yang mereka bimbing. Perhatian yang lebih serius terhadap tugas pekerjaan evaluasi hasil belajar siswa diperlukan sekali, karena mempunyai pengaruh yang berarti terhadap kemajuan pendidikan dan perkembangan mereka dimasa yang akan datang. Perhatian

terhadap pekerjaan guru dalam mengevaluasi hasil belajar siswa yang dimaksud adalah mengetahui sejauhmana guru telah melaksanakan kegiatan evaluasi ini secara benar sesuai dengan kriteria yang ada. Oleh karena itu diperlukan suatu cara untuk menilai kinerja evaluasi guru dalam membuat soal.

Nitko mengemukakan bahwa, penilaian kinerja sering pula disebut dengan penilaian otentik (*authentic assessment*) yakni penilaian yang lebih cenderung melihat proses dan produk secara langsung. Beberapa ciri dari penilaian ini antara lain: (1) menekankan kepada aplikasi (*emphasize application*), (2) berfokus kepada penilaian langsung (*direct assessment*), (3) berhubungan dengan masalah-masalah yang riil (*realistic problems*), dan (4) memanfaatkan gaya berpikir terbuka (*open-ended thinking*). Penilaian kinerja menekankan kepada aplikasi bertujuan untuk melihat sejauhmana seseorang mampu mengintegrasikan aspek-aspek teoritis (dalam hal ini teori-teori evaluasi dan khususnya dalam membuat soal yang bermutu) dengan praktek nyata dalam kehidupan sehari-hari. Berfokus kepada penilaian langsung, bertujuan untuk memperoleh gambaran yang lebih jelas tentang kegiatan yang baru saja dilakukan. Ketika seseorang melakukan sesuatu ketika

itu pula mereka memperoleh skor penilaian tertentu. Berfokus kepada hal-hal yang bersifat riil artinya, memanfaatkan pengalaman yang sebenarnya terjadi, menggambarkan suatu realita alamiah secara apa adanya.

Menurut Mudjijo (1995:61) suatu tes yang baik berarti tes tersebut memiliki butir-butir soal yang baik pula. Pengujian kebaikan suatu tes tidak terlepas dari pengujian atau analisis kebaikan butir-butir soalnya. Analisis butir tes bertujuan untuk meningkatkan validitas dan reliabilitas tes itu secara keseluruhan, sebab validitas dan reliabilitas tes tergantung kepada ciri-ciri butir tesnya.

Analisis butir tes adalah suatu prosedur yang sistematis, yang akan memberikan informasi-informasi yang sangat khusus terhadap butir tes yang telah disusun (Daryanto, 2005:177). Menurut Mudjijo (1995:61), analisis butir tes merupakan suatu tes yang dapat membantu meningkatkan kebaikan suatu tes. Berdasarkan pendapat tersebut dapat dikatakan bahwa sesungguhnya analisis butir tes merupakan suatu kegiatan sistematis yang dilakukan untuk mengetahui butir tes.

Mudjijo (1995:61), mengemukakan bahwa melalui analisis butir tes berarti mencoba menganalisis tingkat kebaikan suatu tes dengan melihat tingkat kebaikan

tiap-tiap butir soalnya. Analisis butir tes ini berfungsi untuk membantu dalam mengidentifikasi butir-butir tes yang jelek, memperoleh informasi yang akan dapat digunakan untuk menyempurnakan tes untuk kepentingan lebih lanjut, dan memperoleh gambaran secara selintas tentang keadaan tes yang kita susun (Daryanto, 2005:177). Melalui analisis butir tes dapat diperoleh informasi tentang kejelekan sebuah tes dan petunjuk untuk mengadakan perbaikan. Dapat disimpulkan bahwa melalui analisis butir tes dapat diketahui ciri dari butir tes tersebut, dengan butir-butir tes yang memenuhi syarat, maka validitas dan reliabilitas tes dapat ditingkatkan. Adapun butir-butir tes yang kurang baik atau tidak memenuhi syarat sebagai butir tes yang baik, baik itu dilihat dari daya pembeda atau taraf sukarnya, hendaknya diperbaiki (direfisi), diganti, atau dibuang. Bila dengan hilangnya butir soal tersebut validitas tes masih dapat dipertanggung jawabkan, maka butir soal itu dapat dibuang, tanpa diganti. Tindakan yang dilakukan terhadap butir-butir tes yang kurang baik, apakah itu direfisi, diganti atau dibuang, tergantung pada kondisi tes itu sendiri, dengan kata lain, kita perlu mempelajari butir tes yang bersangkutan dalam kaitannya dengan tes secara keseluruhan.

Model pengukuran pendidikan untuk menganalisis tes hasil belajar dapat menggunakan model pengukuran model pengukuran klasik dan model pengukuran modern (Dali, 1992:4). Pengukuran klasik mempunyai ciri karakteristik butir tes tidak dapat dilepaskan dari karakteristik peserta. Sebaliknya karakteristik peserta tes tergantung pada karakteristik butir tes. Bagi peserta tes yang memiliki kemampuan rendah butir didasarkan sukar, tetapi butir tes tersebut didasarkan mudah bagi peserta yang memiliki kemampuan tinggi, jadi dalam teori tes klasik bahwa sukar atau mudahnya suatu butir tes tergantung pada tinggi rendahnya kemampuan peserta tes.

Pengukuran modern bertujuan menghilangkan kelemahan yang ada pada teori tes klasik. Teori tes modern berhasil membebaskan saling ketergantungan antara karakteristik butir dengan karakteristik peserta tes (Dali. 1992:160) karakteristik peserta tes tidak lagi ditentukan oleh karakteristik butir tes. Peserta yang memiliki kemampuan yang tinggi akan tetap mempunyai kemampuan yang tinggi, walaupun diuji dengan tes yang memiliki butir tes yang sukar maupun butir tes yang mudah. Sebaliknya karakteristik butir tes tidak lagi ditentukan oleh karakteristik peserta tes. Butir tes yang sukar akan tetap saja sukar, meskipun dijawab oleh peserta tes yang berkemampuan rendah maupun

peserta tes yang berkemampuan tinggi (Fahmi, 2007:30) Ciri yang mendasar dari reori responsi butir ini adalah bahwa kinerja soal dihubungkan dengan sejumlah “sifat laten” responden yang diperkirakan, yang dilambangkan dengan ( $\theta$ ) dibaca theta (Anastasi, 1997:140).

Teori responsi butir dikembangkan atas tiga asumsi dasar yaitu unidimensi, independensi lokal, dan invariansi parameter. Asumsi unidimensi menyatakan bahwa kemampuan yang di ukur adalah kemampuan yang berdimensi tunggal atau hanya satu kemampuan dominan yang menonjol. Asumsi independensi lokal menyatakan bahwa kemampuan peserta tes dibuat konstan, maka jawaban peserta tes terhadap satu butir tidak dipengaruhi jawaban terhadap butir yang lain. Asumsi invariansi parameter menyatakan bahwa karakteristik butir adalah tetap atau tidak berubah sekalipun peserta yang menjawab butir itu berubah-ubah. Karakteristik peserta tidak akan berubah walau butir yang dijawab itu berubah-ubah. Jadi peserta yang berkemampuan tinggi tetap berkemampuan tinggi meski diuji dengan butir tes yang berbeda-beda.

Teori responsi butir menunjukkan model logistik sebagai model teori responsi butir yang sering digunakan, karena prosedur komputasinya lebih mudah dan sederhana. Model logistik adalah model

yang membangun antara karakteristik peserta tes dengan karakteristik butir tes. Hubungan tersebut berbentuk kurva yang disebut kurva karakteristik item (Item Characteristik Curve).

Bentuk lengkungan atau kuva karakteristik butir ditentukan oleh parameter ciri ( $a_i$ ,  $b_i$ ,  $c_i$ ) dan parameter ciri peserta ( $\theta$ ) serta response jawaban peserta terhadap butir dalam bentuk probabilitas jawaban benar  $P_i$  Parameter  $a_i$  adalah ciri butir yang berkenaan dengan daya pembeda yaitu kemampuan butir untuk mempertegas perbedaan di antara peserta yang dapat dan yang tidak dapat menjawab butir itu dengan benar. parameter diskriminasi  $a_i$  mengindikasikan lereng kurva, semakin tinggi nilai  $a_i$  semakin curam lereng kurva. Parameter  $b_i$  adalah ciri butir yang berkenaan dengan faktor terkaan yakni yang dapat menyebabkan peserta secara kebetulan menjawab butir itu dengan benar. Parameter ini menunjukkan probabilitas bahwa sebuah respon yang tepat muncul berdasarkan peluang (Anastasi, 2003:142).

Jumlah parameter butir yang disertakan dalam suatu model akan berpengaruh terhadap bentuk kurva. Dalam teori responsi butir dikenal tiga model yang populer yaitu model logistik satu parameter (L1P) menggunakan parameter taraf sukar butir  $b_i$  model dua parameter (L2P) selai menggunakan parameter taraf sukar butir,

juga menggunakan parameter daya pembeda butir ai, dan model ketiga parameter (L3P) menggunakan semua parameter butir yaitu parameter taraf sukar butir bi, parameter daya pembeda butir ai, dan parameter faktor terkaan ci. Seluruh peserta yang mengerjakan tes itu merupakan populasi peserta. Teori responsi butir menuntut ukuran populasi yang besar. Model 2P menuntut ukuran populasi yang lebih besar dari 1P, dan model 3P menuntut ukuran populasi yang lebih besar dari 2P. Ukuran populasi untuk model 1P menuntut tidak kurang dari 500, untuk model 2P tidak kurang dari 1000, untuk model 3P tidak kurang dari 1500. Bila ukuran populasi kecil akan menghasilkan estimasi parameter yang kurang cermat, makin banyak peserta yang merespon suatu butir tersebut makin stabil nilai yang diperoleh, artinya makin dipercaya hasil yang didapatkan. Pada analisis tes geografi UASBN SMA tahun 2014 ini akan menggunakan model logistik tiga parameter (L3P).

Fungsi informasi pada teori response butir berkaitan secara terbalik dengan ketidakpastian (Dali, 1992:306) ini berarti bahwa makin tinggi ketidakpastian makin rendah fungsi informasi. Sebaliknya makin rendah ketidakpastian makin tinggi nilai fungsi informasi.

Fungsi informasi butir berkaitan dengan ketidakpastian pada butir. Ketidakpastian pada butir berkaitan dengan parameter ciri peserta. Peserta dengan kemampuan tinggi mengerjakan butir yang mudah, maka dengan mudah butir itu dijawabnya dengan benar, sebaliknya peserta dengan kemampuan yang rendah mengerjakan butir yang sukar dengan susah payah maka butir itu masi dijawabnya dengan salah, dalam hal ini kurang diketahui beberapa besar kemampuan peserta adalah tinggi maka fungsi informasi menjadi rendah (Dali,1992:308).

Perangkat uji tes terdiri atas sekelompok butir yang sengaja dikumpul menjadi satu satuan. Dalam hal ini, fungsi informasi tes ditentukan oleh fungsi informasi butir yang membentuk tes itu. Fungsi informasi tes merupakan jumlah dari semua fungsi informasi butir (Dali, 1992:315).

Tes sumatif adalah suatu penilaian yang pelaksanaannya itu dilakukan pada akhir tahun atau akhir program, atau lebih spesifiknya penilaian yang dilakukan pada akhir semester dari akhir tahun. Jadi, tujuannya adalah untuk melihat hasil yang dicapai oleh para siswa, yaitu seberapa jauhkah tujuan-tujuan kurikuler yang berhasil dikuasai oleh para peserta didik, dan penilaian ini pun dititik beratkan pada

penilaian yang berorientasi kepada produk, bukan kepada sebuah proses.

Tes sumatif adalah tes yang dilakukan pada setiap akhir waktu yang di dalamnya tercakup lebih dari satu pokok bahasan, dan dimaksudkan untuk mengetahui sejauhmana peserta didik telah dapat berpindah dari suatu unit ke unit berikutnya. Winkel (1998:39) mendefinisikan tes sumatif sebagai penggunaan tes-tes pada akhir suatu periode pengajaran tertentu, yang meliputi beberapa atau semua unit pelajaran yang diajarkan dalam satu semester, bahkan setelah selesai pembahasan suatu bidang studi.

Penilaian tes sumatif adalah penilaian yang dilakukan untuk memperoleh data atau informasi sampai dimana penguasaan atau pencapaian belajar siswa terhadap bahan pelajaran yang telah dipelajarinya selama jangka waktu tertentu. Adapun fungsi dan tujuannya ialah untuk menentukan apakah dengan nilai yang diperolehnya itu siswa dapat dinyatakan lulus. Pengertian lulus dan tidak lulus disini dapat berarti : dapat tidaknya siswa melanjutkan ke modul berikutnya, dan dapat tidaknya seorang siswa mengikuti pelajaran pada semester berikutnya, dan dapat tidaknya seorang siswa dinaikan ke kelas yang lebih tinggi.

Dari uraian di atas dapat disimpulkan berfungsi dan bertujuan untuk mendapatkan informasi sampai dimana prestasi atau penguasaan dan pencapaian belajar siswa yang selanjutnya diperuntukan dengan penentuan lulus tidaknya seorang siswa, maka penilaian itu disebut penilaian sumatif.

Naga (1992:56) menyatakan bahwa tes yang baik atau tes yang bermutu adalah tes yang dapat memenuhi persyaratan: (1) validitas, (2) reliabilitas, (3) taraf kesukaran, (4) mempunyai daya pembeda dan (5) pengecohnya berfungsi. Mutu soal dapat diketahui dengan cara melakukan analisis soal, bertujuan mengetahui keputusan apa yang tepat diberikan kepada siswa, cara pengajaran, termasuk item soal yang akan diberikan.

Validitas tes adalah kecocokan atau ketepatan suatu tes dalam mengukur sesuatu yang hendak diukur. Crocker dan Algina mengemukakan tiga bentuk validitas tes yaitu: (1) validitas isi (content validity), (2) validitas kriteria (criterionrelated validation) dan (3) validitas konstruk (construct validation). Selanjutnya dijelaskan bahwa validitas isi melihat kesesuaian atau keterwakilan butir dengan keseluruhan materi yang diberikan dalam proses belajar mengajar. Langkah-langkah yang dilakukan untuk mendapatkan validitas isi suatu tes di

antaranya: (1) definisikan kemampuan yang hendak diukur secara jelas, (2) seleksi soal menggunakan panelis.

Reliabilitas (reliability) berarti kekonsistensian atau ketetapan hasil penilaian Nitko dalam Anwar (2009:53). Perhitungan reliabilitas dapat dilakukan untuk tes esay maupun tes objektif. Untuk tes objektif dilakukan dengan metode split half, kuder richardson telah mengembangkan dua rumus untuk mengestimasi reliabilitas internal konsistensi yakni KR 20 dan KR 21.

Tingkat kesukaran diukur melalui sekor butir yang dihasilkan oleh jawaban sejumlah peserta uji tes. Makin banyak peserta uji tes yang berhasil dengan benar mengerjakan suatu butir, makin tidak sukar / mudah butir tersebut. Untuk menentukan taraf sukar butir digunakan rumus:

$$P = \frac{R}{T} \times 100$$

Dimana: P= taraf sukar butir,  
T = peserta tes keseluruhan dan R = jumlah peserta yang menjawab soal dengan benar.

Butir yang baik adalah butir yang tidak terlalu sukar dan tidak terlalu mudah yaitu 0,5. Tuckman dalam Naga menganjurkan bahwa sebaiknya taraf sukar butir itu berada antara  $0,33 \leq P_{tj} \leq 0,67$ . Butir dengan nilai  $P = < 0,33$  adalah butir sukar, sedangkan butir dengan nilai  $P = > 0,67$  adalah butir mudah. Hal ini menunjukkan bahwa, semakin mendekati

angka 1 soal semakin mudah, sebaliknya semakin dekat dengan 0 soal semakin sulit.

Daya beda soal maksudnya, kemampuan butir dengan skornya dapat membedakan responden berkemampuan tinggi dan yang berkemampuan rendah. Gregory (2000) mengemukakan bahwa indek daya beda adalah suatu indek yang menyatakan seberapa efisien suatu item dapat membedakan peserta kelompok tinggi dengan peserta kelompok rendah. Daya beda butir dapat dilihat melalui rumus dimana,  $P_u$  adalah proporsi kelompok atas yang menjawab item dengan benar dan  $P_i$  adalah proporsi kelompok bawah yang menjawab item secara benar dan D nilai indek diskriminasi pada rentangan antara  $-1,00 - 1,00$ . Dengan ketentuan :  $D \geq 0,40$  cukup memuaskan,  $0,30 \leq D \leq 0,39$  = sedikit atau tanpa revisi  $0,20 \leq D \leq 0,29$  = perbatasan atau perlu direvisi  $D \leq 0,19$  dibuang.

## B. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif metoda deskriptif evaluatif. Penelitian deskriptif dalam bidang pendidikan mendeskripsikan fenomena-fenomena kegiatan pendidikan dan pembelajaran. Penelitian deskriptif adalah penelitian yang bertujuan mendeskripsikan secara sistematis, factual dan akurat mengenai dan sifat populasi

tertentu, atau mencoba menggambarkan fenomena secara detail (Lehman, 2007:83 dalam Yusuf). Sedangkan menurut Sukardi (2003:157) penelitian deskriptif merupakan metode penelitian yang berusaha menggambarkan dan menginterpretasikan objek sesuai apa adanya. Penelitian ini dilakukan dengan cara mengumpulkan dan menganalisis data secara sistematis untuk menentukan nilai atau manfaat dari suatu kegiatan pendidikan. Penelitian ini untuk menyempurnakan dan menguji pelaksanaan atau praktek pendidikan. (Nana, 2006:120). Penelitian ini di rancang dengan pendekatan deskriptif kuantitatif karena hasil penelitian ini untuk mendeskripsikan validitas tes, reliabilitas tes, taraf sukar butir dan daya beda butir.

Dalam penelitian ini, peneliti membedakan antara responden dan populasi. Populasi adalah kelompok yang menjadi asal dari sampel yang diambil. Populasi dalam penelitian ini adalah butir-butir soal yang diujikan kepada peserta yang mengikuti ujian semester II Geografi kelas X tahun pelajaran 2013/2014 Kabupaten Padang Pariaman terdiri dari 21 SMA Negeri di Kabupaten Padang Pariaman.

Responden yang mengikuti ujian semester II Geografi kelas X tahun pelajaran 2013/2014 Kabupaten Padang Pariaman terdiri dari 21 SMA Negeri di

Kabupaten Padang Pariaman. Lembar jawaban yang masih tersisa sebanyak 828 lembar terdiri dari 4 sekolah. Data ini merupakan arsip dari guru-guru geografi yang ada di Kabupaten Padang Pariaman.

Teknik pengambilan sampel yang digunakan adalah total sampling, dimana seluruh sambil menjadi sampel, jumlah sampel yang ada adalah 1012 orang. Alasan peneliti mengambil semua responden menjadi sampel karena pada model Logistik Dua Parameter (L2P) pada teori responsi butir jumlah sampel, minimal 1000 peserta, semakin banyak jawaban responden terhadap butir ini semakin baik, karena semakin stabil nilai yang diperoleh. Sehingga dipercaya hasil yang didapatkan.

### **C. HASIL PENELITIAN DAN PEMBAHASAN**

Data yang telah diperoleh akan dianalisis dengan menggunakan teori responsi butir model L2P, sebelumnya dilakukan pengujian persyaratan teori responsi butir yaitu:

#### **1. Pemeriksaan Unidimensi**

Pemeriksaan terhadap adanya dimensi tunggal yang dominan diukur pada tes ujian semester II Geografi kelas X tahun pelajaran 2013/2014 Kabupaten Padang Pariaman dilakukan dengan memakai teknik statistik analisis faktor. Teknik analisis faktor dapat dipakai untuk menganalisis data yang diperoleh dari jawaban tes tes ujian semester I

Geografi kelas X tahun pelajaran 2013/2014 Kabupaten Padang Pariaman, nilai KMO (Kaiser-Meyer-olkin) yang diperoleh cukup besarnya itu 0,700 nilai itu lebih besar dari kriteria 0,600 dengan taraf signifikan uji-barlett's sebesar 0,000 (Fahmi Rizal, Disertasi:2007:116).

## 2. Pemeriksaan Invariansi Parameter

Pemeriksaan invariansi parameter dilakukan untuk melihat apakah parameter peserta tes tidak berubah apabila diuji dengan butir yang memiliki karakteristik berbeda. Untuk itu data dibagi menjadi dua kelompok yaitu data instrument pada butir bernomor ganjil dan data instrument pada butir bernomor genap.

## 3. Pemeriksaan Independensi Lokal

Pemeriksaan independensi lokal bertujuan untuk mengetahui apakah suatu sub populasi kemampuan tes terbebas dari pengaruh karakteristik butir tes. Pemeriksaan independensi lokal dilakukan melalui analisis kovariansi antar interval data skor ujian semester II Geografi kelas X tahun pelajaran 2013/2014 Kabupaten Padang Pariaman.

## 4. Analisis Validitas Tes

Analisis validitas tes dilakukan dengan cara menganalisis jawaban dari peserta tes ujian semester II Geografi

kelas X tahun pelajaran 2013/2014 Kabupaten Padang Pariaman. Berdasarkan analisis tes yang telah dilakukan untuk mengetahui validitas tes ialah dengan menggunakan formula korelasi product moment, maka didapatkan hasil dari 50 butir soal yang valid 33 butir sedangkan yang di drop 17 butir.

## 5. Analisis Reliabilitas

Reliabilitas menunjukkan suatu pada suatu pengertian bahwa suatu instrumen cukup dapat dipercaya untuk dikumpulkan sebagai alat pengumpul data karena instrumen tersebut sudah baik. Mencari reliabilitas dengan KR 20 terdapat pada lampiran 4 dari hasil tersebut maka dapat diperoleh nilai 0,551.

## 6. Analisis Butir Tes Daya Beda dan Taraf Sukar

Analisis butir tes dilakukan dengan cara menganalisis jawaban dari peserta tes ujian semester II Geografi kelas X tahun pelajaran 2013/2014 Kabupaten Padang Pariaman. Data jawaban siswa dianalisis sebanyak 5 kali dengan menggunakan program Ascal Versi 3.2 dari *Assesment System Corporation*. Lalu hasil analisis tersebut dirata-ratakan.

Berdasarkan analisis tes yang telah dilakukan dengan model L2P dengan melihat lengkung slop pada kurva karakteristiknya maka butir soal ujian semester tingkat SMA sekabupaten Padang Pariaman tahun ajaran 2013/2014 dari 50 butir soal yang diterima 33 butir sedangkan yang di tolak 17 butir.

#### **D. SIMPULAN DAN SARAN**

Berdasarkan hasil pembahasan maka dapat ditarik kesimpulan sebagai berikut:

1. Hasil validitas tes ujian semester II mata Geografi kelas X tahun pelajaran 2013/2014 Kabupaten Padang Pariaman terdapat 33 butir soal yang valid dan 17 butir yang drop
2. Hasil reliabilitas tes ujian semester II mata Geografi kelas X tahun pelajaran 2013/2014 Kabupaten Padang Pariaman tidak reliabel dengan 0,551.
3. Hasil tingkat kesukaran butir pelajaran tes ujian semester II mata Geografi kelas X tahun pelajaran 2013/2014 Kabupaten Padang Pariaman berada pada tingkat kesukaran dalam batas diterima.
4. Hasil tingkat daya beda butir tes ujian semester II mata pelajaran Geografi kelas X tahun pelajaran 2013/2014 Kabupaten Padang Pariaman termasuk tinggi dari batasan parameter.

Berdasarkan kesimpulan dan implikasi di atas, maka dikemukakan beberapa saran yaitu:

1. Terhadap butir tes yang di tolak umumnya butir itu telalu mudah dan terlalu sulit, sebaiknya dilakukan perbaikan atau revisi agar butir tersebut sesuai dengan taraf sukar yang distandarkan dalam batasan parameter. Revisi dapat dilakukan dengan memperhatikan konstruksi jawaban atau kalimat soal agar terkesan bahwa butir itu tidak mudah diterka.
2. Untuk butir tes yang mempunyai daya beda yang rendah dilakukan perbaikan sehingga butir tes itu memiliki indeks daya beda yang tinggi dengan memperhatikan karakteristik peserta tes sehingga butir tes yang dibuat benar-benar sesuai dengan karakteristik peserta tes.
3. Evaluasi terhadap butir-butir tes dengan menggunakan teori responsi butir baru pertama kali dilakukan oleh peneliti, oleh sebab itu disarankan kepada peneliti lain untuk melanjutkan penelitian ini agar penggunaan teori modern ini lebih berkembang dan lebih diminati dalam dunia pendidikan.

## DAFTAR PUSTAKA

- Anastasi, Anne dan Susan. 1997. *Tes Psikologi*. Terjemahan oleh Robertus H. Imam. 2003. Jakarta : Gramedia.
- Anwar, Syafri. 2009. *Peningkatan Kinerja Evaluasi Melalui Feed Back Tes Buatan Guru*, Padang (Disertasi).
- Arikunto, Suharsimi. 1997. *Penilaian Program Pendidikan*. Jakarta: Depdikud-Direjen Dikti.
- Azwar, Saifuddin. 2000. *Reliabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar.
- Bloom, Benjamin.S. 1976. *Human Characteristic and School Learning*. New York: McGraw –Hill Book Company.
- Dali S, Naga. 1992. *Pengantar Teori Sekor Pada Pengukuran Pendidikan*. Jakarta: Gunadarma Press.
- Dalin, Per. 1998. *School Development Theories and Strategies*. New York: Redwood books.
- Daryanto H. 2005. *Evaluasi Pendidikan*. Jakarta: PT. Rineka Cipta.
- \_\_\_\_\_. 1999. *Evaluasi Pendidikan*. Jakarta: PT. Rineka Cipta.
- Depdiknas. 2003. *Kebijakan Umum Program Pengembangan Pendidikan Menengah Umum*. Jakarta: Direktorat Dikmenum.
- Dick, Walter & Reiser Robert A. 1996. *Planning Effective Instruction*. Needham: Allyn and Bacon.
- Fahmi, Rizal. 2007. *Penerapan Teknik Penghalusan Taksiran Kemampuan Peserta Tes Dilihat Dari Kestabilan Koefisien Penyetaraan*. Disertasi tidak Diterbitkan. Jakarta: Program Pascasarjana UNJ.
- Gentry, Castelle.G. 1994. *Introduction to Instructional Development*. Process and technique. California: Wadsworth Publishing Company.
- Gordon, Thomas. T.T.T, 1997. *Teacher Effectiveness Training*, MGE, Menjadi Guru Efektif. Jakarta : PT. Gramedia Pustaka Utama.
- Gregory, Robert L. 2000. *Psychological Testing, History, Principle, and Applications*, Needham Height.
- Gronlund, Norman E. 1990. *Measurement and Evaluation in Teaching*, 5<sup>th</sup> edition, New York: MacMillan Publishing Company.
- Hill, B. Campbell & Ruptic, Cynthia A. 1994 *Practical Aspects of Authentic Assessment*, Washington : Cristoper Gordon, Inc.
- Kemp, Jerrold E., Marrison, G.R., Ross S.M., 1985. *Designing Effective Instruction*. NewYork: Macmillan College Publishing Company, Inc.

- Linn, Robert L. and Gronlund, Norman E. 1995. *Measurement and Assessment In Teaching*. Ohio: Prentice-Hall, Inc.
- Little, Angela, & Wolf, Allison. 1996 *Assessment in Transition, Learning Monitoring and selection in International Perspective*. Pargamon: Eksevier Science Ltd.
- McBeath, R.J. 1992. *Instructing and Evaluating in High Education: Educational Technology A guide Book For Planning Learning out Comes*, New Jersey: Publication, Inc. Englewood Cliffs.
- Medley,DM., Coker,H., & Soar,RS. 1984. *Measurement Based Evaluation of Teacher Performance, An Emperical Approach*, New York & London: Longman Inc.
- Meilanne. 2009. *Evaluasi Kualitas Tes Bahasa Indonesia dan Matematika UASBN Sekolah Dasar Tahun 2008 Di Kota Padang Berdasarkan Model L2P Teori Responsi Butir*. *Tesis* tidak Dipublikasikan. Padang: Program Pascasarjana UNP.
- Mudjiyo, 1995. *Tes Hasil Belajar*. Jakarta: Bumi Aksara
- Mulyana, 2009. *Evaluasi Kualitas Tes Matematika Ujian Akhir Sekolah Berstandar Nasional SD Tahun 2008 di Kota Padang berdasarkan Analisis Teori Responsi Butir*. *Tesis* tidak Dipublikasikan. Padang: Program Pascasarjana UNP
- Naga, Dali S., *Pengantar Teori Sekor pada Pengukuran Pendidikan*, Jakarta: Penerbit Gunadarma, 1992.
- Nana, Syaodih Sukmadinata. 2006. *Metode Penelitian Pendidikan*. Bandung: Remaja Rosdakarya.
- Nitko, Anthony J. 1996. *Educational Assessment of Studies*, 2<sup>nd</sup> edition, New Jersey: Prentice-Hall, Inc.
- Rosmalinda, 2013. *Analisis Kualitas Tes Pra UN Geografi SMA/MA Tingkat Kabupaten Dharmasraya Tahun Ajaran 2012/2013. Berdasarkan Model Logistik Satu Parameter (L1P) Teori Responsi Butir*. *Tesis* tidak Dipublikasikan. Padang: Program Pascasarjana UNP.
- Saifuddin, Azwar. 2003. *Tes Prestasi: Fungsi dan Pengembangan Pengukuran Prestasi Belajar*. Yogyakarta: Pustaka Belajar.
- Salinan Undang-Undang Republik Indonesia No. 20 tahun 2003 tentang Sistem Pendidikan Nasional.
- Soedijarto. 2000. *Pendidikan Nasional, Sebagai Wahana Mencerdaskan Kehidupan Bangsa dan Membangun Peradaban Bangsa*. Jakarta: Cinaps.

Sukardi. 2003. *Metodologi Penelitian Pendidikan Kompetensi dan prakteknya*. Jakarta: Bumi Aksara.

Sumadi, Suryabrata. 1984. *Pengembangan Tes Hasil Belajar*. Jakarta: Rajawali Pers.

Suryabrata, Sumadi. 2002. *Pengembangan alat ukur psikologis*, Yogyakarta: Penerbit Andi.

Wiersma William & Jurs, G. Stephen. 1990. *Educational Measurement And Testing, 2<sup>nd</sup> Edition*, Boston: Allyn and Bacon.

Yusuf. 2000. *Evaluasi Program*. Jakarta: Penerbit Rineka Cipta.