

## **ESTIMASI KESALAHAN BAKU PENGUKURAN SOAL UAS MATEMATIKA SMAN DI KOTASELATPANJANG KABUPATEN KEPULAUAN MERANTI**

Dedek Andrian, Djemari Mardapi  
Prodi Penelitian dan Evaluasi Pendidikan PPs UNY, Universitas Negeri Yogyakarta  
dedek\_pendidikan@yahoo.co.id, djemarimardapi@gmail.com

### **Abstrak**

Penelitian ini bertujuan: (1) mengestimasi kesalahan pengukuran soal-soal UAS Mata Pelajaran Matematika SMAN Jurusan IPA dan IPS yang dirancang oleh guru SMAN 1, SMAN 2, SMAN 3, MAN; dan (2) melihat akurasi berbagai metode dalam mengestimasi kesalahan baku pengukuran. Analisis data didasarkan pada respon peserta didik terhadap perangkat tes UAS kelas XII Mata Pelajaran Matematika tahun ajaran 2013/2014. Sumber data berupa lembar jawaban dari 288 siswa SMAN 1, 215 dari SMAN 2, 103 dari SMAN 3, dan 101 dari MAN. Analisis butir soal perangkat tes tersebut dilakukan dengan Metode Thorndike, Metode *compound binomial*, Metode Analisis Varians, dan Metode Teori Respons Butir. Hasil analisis estimasi kesalahan baku pengukuran berdasarkan Metode Thorndike terkecil terdapat pada perangkat tes SMAN 1 Jurusan IPA, sedangkan nilai terbesar terdapat pada perangkat tes SMAN 3 Jurusan IPA. Berdasarkan Metode *compound binomial* nilai estimasi kesalahan pengukuran terkecil terdapat pada perangkat tes MAN Jurusan IPS, sedangkan nilai terbesar terdapat pada perangkat tes SMAN 3 Jurusan IPA. Berdasarkan Metode Analisis Varians nilai estimasi kesalahan baku pengukuran terkecil terdapat pada perangkat tes SMAN 3 Jurusan IPS, sedangkan nilai terbesar terdapat pada perangkat tes SMAN 1 Jurusan IPS. Berdasarkan Metode Teori Respons Butir nilai estimasi terkecil terdapat pada Model Logistik 1-Parameter pada perangkat tes MAN Jurusan IPA, yaitu pada  $\theta=0$  dengan nilai sebesar 0,322, sedangkan nilai estimasi terbesar terdapat pada Model Logistik 3-Parameter pada perangkat tes MAN Jurusan IPS, yaitu pada  $\theta=-3$  dengan nilai sebesar 2,232. Akurasi Metode estimasi kesalahan baku pengukuran secara berturut-turut adalah Metode *compound binomial*, Metode Thorndike, Metode Analisis Varians dan Metode Teori Respons Butir, yaitu 0,917, 1,953, 2,108 dan 2,232.

**Kata kunci:** *estimasi, kesalahan pengukuran, perangkat tes.*

## **ESTIMATION STANDARD ERROR OF MEASUREMENT IN MATHEMATICS TEST ITEMS OF SENIOR HIGH SCHOOLS IN SELATPANJANG, KEPULAUAN MERANTI DISTRICT**

Dedek Andrian, Djemari Mardapi  
Prodi Penelitian dan Evaluasi Pendidikan PPs UNY, Universitas Negeri Yogyakarta  
dedek\_pendidikan@yahoo.co.id, djemarimardapi@gmail.com

### **Abstract**

*This study aims to: (1) estimate the standard error of measurement of year 12 mathematics subject in high school natural science and social science program, designed by teachers of SMAN 1, SMAN 2, SMAN 3, MAN; and (2) investigate the accuracy of the various methods in estimating the standard error measurement. The data analysis was based on students' responses to End-of-Semester Test Items in the academic years 2013/2014. The data source was computerised answer sheets from 288 students' of SMAN 1, 201 of SMAN2, 103 of SMAN 3, and 101 of MAN. The analysis on the test items of eight test instrument was done using the Thorndike Method, compound binomial Method, ANAVA Method, and Item Response Theory Method. The results of the analysis of the estimated SEM based on the estimated value the Thorndike Method, show the smallest SEM is found in the test item SMAN 1 natural science program, while the largest value is the test item SMAN 3 natural science program. Based on the estimated value of the compound binomial Method, the smallest SEM is found in the test item MAN social science program, while the largest value is found in the test item SMAN 3 natural science program. Based on the Varians Analysis Method, the smallest SEM is found in the test item SMAN 3 social science program, while the largest value is found in the test item SMAN 1 social science program. Based on the Item Response Theory Method, the smallest SEM is found in the 1 - Parameter Logistic Model on the test item MAN natural science program, namely  $\theta = 0$  with the value of 0.322, while the largest estimated value is found in the 3 - Parameter Logistic Model on the test item MAN social science program, namely at  $\theta = -3$  with the value of 2.232. Accuracy of the estimated SEM methods respectively is compound binomial Method, Thorndike Method, Varians Analysis Method, and Item Response Theory Method, namely 0.917, 1.953, 2.108 and 2.232.*

**Keywords:** *estimation, measurement error, test instrument.*

## Pendahuluan

Keberhasilan dalam proses pembelajaran dapat dilihat melalui proses penilaian. Penilaian akan menyediakan informasi secara komprehensif tentang keberhasilan dalam proses pembelajaran. Tinggi atau rendahnya hasil belajar siswa akan diinterpretasi melalui proses penilaian dengan standar yang telah ditentukan. Tinggi rendahnya hasil belajar siswa yang dihasilkan melalui penilaian selanjutnya akan dievaluasi sebagai sebuah *judgment* terhadap pembelajaran yang dilaksanakan dalam kurun waktu tertentu. Proses penilaian terhadap keberhasilan pembelajaran tersebut tentunya tidak terlepas dari proses pengukuran yang dilaksanakan dengan instrumen tertentu. Hasil pengukuran yang dilakukan dengan instrumen tertentu itulah yang selanjutnya dinilai dan dievaluasi.

Pengukuran pada dasarnya merupakan penentuan angka bagi suatu objek secara sistematis. Penentuan angka ini merupakan usaha untuk menggambarkan karakteristik suatu objek (Mardapi, 2012, p.7). Senada pendapat tersebut Allen & Yen (1979, p. 2) menyatakan bahwa "*Measurement is the assigning of numbers to individuals in a systematic way as a means of representing properties of the individuals*". Pengertian tersebut dapat diartikan pengukuran adalah penilaian dari number pada individu dengan cara sistematis sebagai sebuah alat untuk mewakili sifat dari individu.

Gronlund & Linn (1990, p. 5) menyatakan bahwa "measurement is the process of obtaining a numerical description of the degree to which an individual possesses a particular characteristic". Makna tersebut dapat diartikan bahwa pengukuran merupakan proses untuk memperoleh sebuah gambaran secara angka tentang sejauh mana tingkat kemajuan individu dalam karakteristik tertentu. Senada pendapat tersebut Miller (2008, p.1) menyatakan bahwa "measurement is a quantitative description of individual's achievement on a single test or multiple assessments". Makna tersebut juga dapat diartikan bahwa pengukuran merupakan deskripsi kuantitatif tentang prestasi individu pada sebuah tes atau beberapa penilaian.

Berdasarkan pernyataan-pernyataan para ahli tersebut dapat disimpulkan bahwa pengukuran merupakan proses menterjemahkan/penentuan kemampuan, bakat, prestasi yang dimiliki siswa dari proses pengukuran dengan sebuah angka berdasarkan sebuah

proses pengukuran atau pengujian. Penentuan angka dalam proses pengukuran memiliki dasar dan kaidah tertentu tergantung skala apa dan alat ukur apa yang digunakan oleh lembaga pendidikan atau guru dalam mengukur kemampuan peserta didik. Apabila guru ingin menilai kemampuan kognitif siswa tentang penguasaan terhadap materi yang telah diajarkan, guru bisa membuat tes tertulis sesuai dengan materi yang telah diajarkan (kompetensi dasar dan standar kompetensi). Apabila guru ingin menilai afektif atau sifat peserta didik, guru bisa membuat catatan kualitatif, lembar observasi atau mengembangkan instrumen yang sifatnya mengarah kepada catatan kualitatif. Begitu juga jika guru ingin menilai aspek psikomotor peserta didik, guru bisa membuat instrumen non tes yang dalamnya terdapat skala-skala sendiri yang bisa ditransformasikan ke dalam angka.

*Assessment* merupakan istilah umum yang didefinisikan sebagai sebuah proses yang ditempuh untuk mendapatkan informasi yang digunakan dalam rangka membuat keputusan-keputusan mengenai para siswa, kurikulum, program-program, dan kebijakan pendidikan, metode atau instrumen pendidikan lainnya oleh suatu badan, lembaga, organisasi, atau institusi resmi yang menyelenggarakan suatu aktivitas tertentu (Uno & Koni, 2012, p.1). Senada dengan pendapat tersebut Miller, Linn, & Gronlund (2009, p.28) menyatakan bahwa "*assessment is a general term that includes the full range of procedures used to gain information about student learning (observation, ratings of performances or projects, paper-and-pencil tests) and the formation for value judgment concerning learning progress*". Pernyataan tersebut dapat diartikan bahwa penilaian merupakan istilah umum yang mencakup prosedur yang digunakan untuk memperoleh informasi tentang belajar siswa (observasi, tingkatan dari kinerja atau proyek, dan tes tertulis) dan pembentukan untuk menilai keputusan mengenai kemajuan pembelajaran.

Phye (1997, p.9) menyatakan bahwa terdapat paling sedikit empat defenisi dari penilaian yang bisa dilihat diliteratur sekarang yang berhubungan dengan penilaian untuk beberapa pendidik (*educator*).

1. *Assessment refers to new formats for gathering information about student's achievement. For example " portfolio assessment"*.

2. *Assessment refers to a new toward gathering information, an attitude that is perhaps kinder and gentler than that represented by standardized testing.*
3. *Assessment has also come to represent a new ethos, one of empowerment, in which assessment are designed and implemented primarily to serve the information needs of student and teacher.*
4. *Assessment has been used to refers to a new process.*

Berdasarkan pernyataan tersebut dapat disimpulkan bahwa penilaian merupakan sebuah format untuk mengumpulkan informasi tentang prestasi siswa, penilaian bisa digunakan untuk melihat prestasi, sikap dan etos.

Penilaian atau *assessment* merupakan komponen penting dalam penyelenggaraan pendidikan. Penilaian yang dilakukan dalam pembelajaran merupakan bagian dari peningkatan kualitas pendidikan. Menurut Rasid & Mansur (2008, p.7) upaya peningkatan kualitas pendidikan dapat ditempuh melalui peningkatan kualitas pembelajaran dan kualitas sistem penilaiannya, oleh karena itu guru memiliki tanggung jawab yang besar dalam proses pembelajaran dan penilaian. Informasi dari keberhasilan pembelajaran yang dilaksanakan melalui pengukuran tidak bisa diperoleh tanpa penilaian.

Wright (2008,p.5) menyatakan bahwa *“Assessment provides an accounting of how much children learn in school and what resources are expended on achieving those learning outcomes”*. Pernyataan tersebut dapat diartikan penilaian menyediakan sebuah kumpulan data tentang bagaimana anak belajar di sekolah dan sumber apa yang telah dikeluarkan untuk mencapai prestasi hasil belajar. Berdasarkan pernyataan tersebut dapat disimpulkan bahwa, guru harus memiliki kompetensi yang baik dalam melaksanakan proses penilaian dan mampu meningkatkan sistem penilaian di sekolah agar bisa mengumpulkan data yang akurat tentang keberhasilan belajar siswa.

Penilaian merupakan bagian dari pertanggung jawaban guru yang sangat dibutuhkan dalam pembelajaran. Perkembangan siswa di sekolah harus diukur untuk melihat kompetensi-kompetensi apa saja yang dimiliki siswa yang telah didapatkan siswa melalui proses pembelajaran. Wright (2008, p.5) menyatakan bahwa *“Educational accountability requires that all student be assessed to quantify that they have learned and what skills they have developed”*.

Pengertian tersebut dapat diartikan bahwa pertanggung jawaban pendidikan dalam pembelajaran dibutuhkan, karena semua murid dinilai untuk mengukur apa yang telah diajarkan dan skill apa yang dikembangkan melalui pembelajaran. Berdasarkan pernyataan tersebut dapat disimpulkan bahwa penilaian yang dilakukan oleh satuan pendidikan merupakan bagian dari akuntabilitas sekolah untuk memperlihatkan sejauh mana perkembangan peserta didik dalam proses pembelajaran. Akuntabilitas kepada orang tua, kepada sekolah maupun kepada pemerintah.

Penilaian merupakan hal yang penting dalam proses pembelajaran. Penilaian merupakan sebuah *support* untuk melihat tentang sejauh mana keberhasilan siswa dalam pembelajaran. Gardner (2012, p.11) menyatakan bahwa *“assessment in education must, first and foremost, serve the purpose of supporting learning”*. Senada pendapat tersebut Campbell & Levin (2009, p.47) menyatakan bahwa *“Assessment for Learning is an important strategy for connecting instructional strategies and classroom practices to individual needs, progress and learning outcomes of student”*. Berdasarkan pernyataan tersebut dapat disimpulkan bahwa penilaian dalam pembelajaran merupakan strategi penting untuk menghubungkan strategi pembelajaran dan praktik kelas untuk kebutuhan individu, kemajuan, dan hasil belajar siswa.

Penilaian dalam pembelajaran harusnya dilakukan secara kontiniu, hal tersebut dilakukan agar kelemahan dan kekurangan dalam pembelajaran dapat disadari secepat mungkin. Terdapat dua penilaian dalam pembelajaran, yaitu penilaian formatif dan penilaian sumatif. Heritage (2010,p.18) menyatakan bahwa,

*Formative Assessment helps teachers adapt their instruction to meet students' needs and assists students to determine what learning adjustment they need to make. Sumative Assessment helps determine whether a student has reached a certain level of competency “ after competing a particular phase of education. Sumatif assessment are used to measure what students have learned. Formatif assessment refer to frequent, interactive assessment of student progress and understanding to identify learning need and adjust teaching appropriately.*

Berdasarkan pernyataan tersebut dapat disimpulkan bahwa penilaian formatif dilakukan untuk membantu guru dalam menyesuaikan pembelajarannya, untuk memenuhi kebutuhan siswa, dan membantu siswa dalam menentukan pembelajaran yang sesuai dibutuhkan oleh siswa. Penilaian sumatif dilakukan untuk menentukan apakah siswa telah mencapai sebuah kompetensi pada level tertentu setelah menyelesaikan tahap-tahap tertentu dari pembelajaran dalam satuan pendidikan.

Evaluasi ialah proses yang menentukan sampai sejauh mana tujuan pendidikan dapat dicapai (Tayibnapis, 2008, p.3). Menurut Worthen & Sander (1973, p.19) "*Evaluation is the determination of the worth of a thing. It include obtaining information for use in judging the worth of a program, product, procedure, or objectif or the potential utility of alternative approaches designed to attain specified objectives*". Makna tersebut dapat diartikan bahwa evaluasi adalah penentuan sebuah nilai dari sebuah benda, ia termasuk memperoleh informasi untuk digunakan dalam penilaian hasil dari sebuah program, pelaksanaan, prosedur, objektif, atau kegunaan secara potensial dari pendekatan alternatif yang telah didesain untuk mencapai hasil yang ditetapkan.

Senada pendapat Gronlund & Linn (1990, p.5) menyatakan bahwa "*evaluation is the sistematic process of collecting, analyzing, and interpreting information to determine the extent to which pupils are achieving intructional objectives*". Berdasarkan makna tersebut dapat disimpulkan bahwa evaluasi merupakan sebuah proses sistematis untuk mengumpulkan, menganalisis, menginterpretasi informasi dari proses pengukuran dan penilaian untuk menentukan sejauh mana siswa telah mencapai tujuan dari pembelajaran.

Evaluasi adalah sebuah riset untuk mengumpulkan, menganalisis dan menyajikan informasi yang bermanfaat mengenai objek evaluasi, menilainya dengan membandingkannya dengan indikator evaluasi dan hasilnya dipergunakan untuk mengambil keputusan mengenai objek evaluasi (Wirawan, 2011, p.7). Makna tersebut senada dengan pendapat Stufflebeam (1971) dalam Mehrens & Lehmann (1991, p.4) menyatakan bahwa "*evaluaition is process of delineating, obtaining, and providing useful information for judging decision alternatives*". Pernyataan tersebut dapat diartikan evaluasi adalah sebuah proses menggambarkan, memperoleh dan menyediakan informasi yang

berguna untuk menilai alternatif keputusan. Berdasarkan pernyataan tersebut dapat disimpulkan bahwa evaluasi tidak akan bisa berlangsung tanpa adanya pengukuran dan penilaian, sehingga bisa dideskripsikan bahwa fungsi pengukuran, penilaian, dan evaluasi merupakan suatu fungsi yang linear, di mana pengukuran proses atau prosedur penetapan angka-angka dengan cara yang sistematis dengan menjaringnya melalui sebuah tes, penilaian proses membandingkan antara angka-angka yang telah didapatkan melalui pengukuran dan evaluasi merupakan langkah perbaikan ketika terjadi kesenjangan dalam proses penilaian.

Terdapat dua bentuk evaluasi dalam pembelajaran di dunia pendidikan, yaitu evaluasi formatif dan evaluasi sumatif. Evaluasi sumatif merupakan evaluasi yang dilakukan setelah menyelesaikan sub pokok pembelajaran atau kompetensi dasar pada mata pelajaran tertentu dalam bentuk ulangan harian, sedangkan evaluasi sumatif merupakan evaluasi yang dilakukan diakhir keseluruhan pembelajaran atau akhir semester. Menurut Miller (2008, p.5) evaluasi formatif dan sumatif adalah sebagai berikut,

*Formatif evaluation is used to diagnose individuals' strengths (competencies that student mastered) and (cognitive, affective, and psychomotor inadequacies) before intruction in an effort to prepare appropriate instructional strategies. Sumatif evaluation takes place after intruction and is used to evaluasi student achievement and overall of the intructional process.*

Berdasarkan pernyataan tersebut dapat disimpulkan bahwa evaluasi formatif merupakan evaluasi yang digunakan untuk melihat penguasaan pembelajaran siswa pada materi tertentu, sedangkan evaluasi sumatif merupakan evaluasi diakhir keseluruhan pembelajaran untuk melihat sejauhmana penguasaan seluruh pembelajaran siswa dan menentukan prestasi siswa pada pembelajaran dalam satu semester.

Penerapan evaluasi formatif dan sumatif dalam pembelajaran pada satuan pendidikan tidak selalu digunakan secara bersamaan. Tidak jarang beberapa sekolah hanya menerapkan evaluasi sumatif untuk mengevaluasi penguasaan dari keseluruhan pembelajaran dan terdapat juga beberapa sekolah menerapkan kedua evaluasi tersebut dalam pembelajaran. Hal yang terbaik dilakukan dalam satuan

pendidikan adalah melakukan evaluasi formatif. Hal tersebut senada dengan penelitian yang telah dilakukan oleh Wiliam (2006, p.288) yang menyimpulkan bahwa evaluasi formatif sangat berguna dalam pembelajaran untuk menghasilkan bimbingan yang sempurna (*counsel of perfection*) dalam pembelajaran.

Pengukuran dalam pembelajaran merupakan kegiatan melakukan kuantifikasi gejala atau objek. Kuantifikasi dalam pengertian ini dapat diartikan suatu kegiatan menterjemahkan data yang diperoleh melalui pengukuran pada objek atau gejala tertentu dengan sebuah angka (Mardapi, 2012, p.1). Gejala atau objek tersebut bisa berupa motivasi prestasi, percaya diri atau prestasi yang semuanya dinyatakan dalam bentuk angka. Angka-angka yang telah didapatkan melalui pengukuran, memberikan informasi kepada guru tentang ketercapaian dari proses belajar yang telah berlangsung. Tentunya informasi yang didapatkan melalui pengukuran adalah hasil dari proses analisis data dengan teknik tertentu. Selanjutnya barulah guru melakukan penilaian secara intensif untuk melihat keberhasilan dari proses pembelajaran, apakah angka-angka yang didapatkan melalui pengukuran sudah memenuhi standar keberhasilan atau sebaliknya.

Untuk Melakukan pengukuran dibutuhkan sebuah instrumen yang baik. Instrumen yang baik adalah instrumen yang memiliki validitas dan reliabilitas tinggi serta memiliki error yang sekecil mungkin dalam menjangkau informasi tentang keberhasilan proses pembelajaran. Instrumen yang baik akan menghasilkan pengukuran yang akurat dan cermat dalam menggali informasi tentang keberhasilan proses pembelajaran (Azwar, 2013, p.173). Senada dengan pendapat tersebut Mardapi (2007, p.67) menyatakan untuk menghasilkan informasi yang akurat maka instrumen dalam pengukuran harus handal, sehingga bisa menghasilkan kesalahan pengukuran yang sekecil mungkin.

Hal yang sulit dirasakan oleh guru dalam pembelajaran adalah membuat instrumen untuk melakukan pengukuran. Banyak guru yang belum terampil dalam membuat alat ukur yang valid dan reliabel. Hal tersebut terbukti berdasarkan penelitian yang dilakukan oleh Widiastuti (2008) yang menyimpulkan bahwa banyak soal-soal ujian sekolah yang dibuat guru tidak layak untuk diujikan dalam proses pengukuran. Hal yang sama dikatakan oleh Feld (1995, p. 295) menyatakan bahwa reliabilitas tes yang dirancang oleh guru (*educator*) relatif

lemah. Berdasarkan pernyataan di atas dapat disimpulkan bahwa untuk membuat alat ukur yang baik bukanlah hal yang mudah, oleh karena itu guru harus menyadari kelemahan dan kekurangannya dalam merancang soal dan berusaha untuk memperbaiki kelemahan dan kekurangan tersebut.

Salah satu upaya yang bisa guru lakukan agar butir soal yang dirancangnya memiliki validitas dan reliabilitas tinggi adalah dengan cara menganalisis butir soal pengukuran secara kontinu. Hasil analisis tersebut akan memberikan informasi kepada guru tentang kelayakan atau validitas butir soal yang digunakan dalam proses pengukuran. Butir soal yang layak atau valid bisa dijadikan guru sebagai bank soal, sedangkan butir soal yang tidak layak digunakan harus diganti atau direvisi untuk diujikan lagi pada pengukuran berikutnya. Selain validitas, hasil analisis juga memberikan informasi kepada guru tentang reliabilitas soal yang dirancang guru. Informasi tentang reliabilitas soal juga penting diketahui oleh guru. Hal tersebut dikarenakan butir soal dikatakan baik apabila telah memenuhi kedua syarat tersebut. Meskipun pada dasarnya butir soal yang memiliki validitas tinggi pada umumnya reliabel, akan tetapi pengujian reliabilitas perlu dilakukan untuk melihat keajegan/kehandalan soal tersebut dalam menjangkau informasi keberhasilan siswa dalam proses pembelajaran.

Pengetahuan dalam membuat butir soal yang baik bagi guru merupakan harga mati dalam proses pengukuran. Guru harus menyadari pentingnya sebuah instrumen yang baik dalam melakukan proses pengukuran. Guru harus berusaha keras, belajar, mengikuti pelatihan-pelatihan dalam rangka meningkatkan kompetensi guru dalam membuat soal yang baik agar soal yang dirancangnya memiliki validitas dan reliabilitas yang tinggi. Soal yang valid dalam pengukuran merupakan soal yang memiliki akurasi dan sah dalam menjangkau informasi tentang keberhasilan proses pembelajaran, sedangkan soal reliabel merupakan soal yang ajeg ketika digunakan berkali-kali dalam kurun waktu tertentu. Ketika guru sudah mampu menghasilkan instrumen yang valid dan reliabel, maka tingkat kesalahan dalam pengukuran dapat dipastikan akan relatif kecil dan guru tersebut bisa melihat kemampuan sebenarnya yang dimiliki oleh siswa.

Tes merupakan sebuah alat atau prosedur yang digunakan untuk mengetahui atau mengukur sesuatu dalam suasana, dengan cara

dan aturan-aturan yang sudah ditentukan (Suharsimi, 2012, p. 67). Senada pendapat tersebut Reynolds, Livingston, & Wilson (2010, p.3) menyatakan bahwa "*a test is a device or procedure in which a sample of an individual's behavior is obtained, evaluated, and scored using standardized procedures*". Pengertian tersebut dapat diartikan, tes adalah sebuah alat atau prosedur di mana sebuah sampel dari perilaku individu diperoleh, dievaluasi, dan diskor dengan menggunakan prosedur yang telah distandarkan.

Senada dengan pendapat tersebut Allen & Yen (1979, p.1) menyatakan bahwa "*A test is a device for obtaining a sample of an individual's behavior*". Berdasarkan pengertian tersebut dapat disimpulkan bahwa tes merupakan salah satu alat untuk melakukan pengukuran, yaitu alat untuk mengumpulkan informasi karakteristik suatu objek. Mengumpulkan informasi bagaimana hasil dari proses pembelajaran yang telah berlangsung dalam satuan pendidikan, baik di SD, SMP, SMA, SMK maupun di Perguruan Tinggi. Melalui tes kemampuan seseorang siswa dapat diketahui secara langsung berdasarkan pembelajaran yang telah dilaksanakan, yaitu melalui respons seseorang terhadap stimulus atau pertanyaan yang diberikan kepada siswa yang mengikuti tes. Tes akan memberikan informasi tentang prestasi belajar, pengetahuan, kemampuan, bakat, dan kepribadian seseorang melalui ujian tertulis, lisan, atau wawancara.

Tes terdiri atas sejumlah pertanyaan yang memiliki jawaban yang benar atau salah, atau semua benar atau sebagian benar. Tujuan melakukan tes adalah untuk mengetahui pencapaian belajar atau kompetensi yang telah dicapai peserta didik untuk bidang tertentu. Senada dengan pendapat tersebut Gronlund (1977, p.19) menyatakan bahwa "*tests can be used in an instructional program to assess entry behavior, monitor learning progress, diagnose learning difficulties and measure performance at the end of instruction*". Berdasarkan pernyataan tersebut dapat disimpulkan bahwa tes bertujuan untuk melihat kemajuan hasil belajar, untuk mendiagnosa kesulitan belajar dan mengukur kinerja pada proses pembelajaran.

Sebagai seorang tenaga pendidik, guru seyogyanya sebelum membuat, mengembangkan dan melaksanakan tes perlu memahami alur pembelajaran dengan baik dan benar. Menyelenggarakan penilaian proses dan hasil belajar

merupakan kompetensi pedagogik yang harus dikuasai oleh guru. Kenyataan yang sering dihadapi di sekolah dan di kampus, seorang guru atau dosen yang telah memahami dan menjiwai proses pembelajaran, dan sering mengikuti pelatihan-pelatihan dalam meningkatkan kemampuan pedagogik, mampu menghasilkan pembelajaran yang bermakna dan mampu merancang tes dengan baik ( Rasid & Mansur, 2008, p. 175). Untuk membuat soal yang baik yang memiliki validitas dan reliabilitas tinggi merupakan hal yang sulit, tapi itu bukanlah hal yang tidak mungkin untuk dilakukan. Seorang guru harus bekerja keras dan mau mengikuti pelatihan-pelatihan yang berkaitan dengan peningkatan kompetensi guru dalam membuat soal. Gronlund (1977, p.18) memberikan petunjuk bagaimana merancang soal dengan baik antara lain: (1) tentukan tujuan dari tes; (2) Identifikasi hasil belajar yang ingin diukur oleh tes; (3) mendefenisi hasil belajar secara spesifik; (4) garis besar materi pelajaran yang akan diukur dengan tes; (5) menyiapkan tabel spesifikasi; (6) menggunakan tabel spesifikasi sebagai dasar untuk mempersiapkan tes.

Berdasarkan keterangan tersebut dapat disimpulkan bahwa memahami cara bagaimana mengembangkan sebuah alat ukur baik itu tes maupun non tes merupakan salah satu kompetensi yang harus dimiliki oleh guru. Jika guru telah memiliki kompetensi yang baik dalam membuat, mengembangkan, merancang dan melaksanakan tes, maka dapat dipastikan alat ukur yang dirancang oleh guru berkemungkinan besar valid, handal dan memiliki kesalahan baku pengukuran yang relatif kecil ketika alat ukur tersebut digunakan.

Pengukuran, Penilaian, dan evaluasi merupakan kegiatan tidak bisa dipisahkan. Dalam kegiatan tersebut akan menghasilkan kesalahan baku dalam pengukuran. Terdapat dua kategori kesalahan dalam pengukuran yaitu kesalahan pengukuran acak dan kesalahan pengukuran secara sistemik (Mardapi, 2008, p.68). Kesalahan secara acak disebabkan kesalahan dalam menentukan sampel isi tes, variasi emosi, lingkungan belajar yang kurang kondusif, kondisi tubuh yang kurang fit dan waktu pelaksanaan tes itu sendiri yang mempengaruhi kinerja siswa dalam menjawab soal serta subjektivitas guru dalam memberi skor. Kesalahan pengukuran sistematik adalah kesalahan pengukuran yang disebabkan oleh soal tes terlalu mudah atau terlalu sukar. Terdapat pendidik yang cenderung memberika soal yang

sulit dan beberapa guru cenderung memberikan soal yang mudah. Selain itu terdapat beberapa guru yang terlalu murah dalam memberi skor dan terlalu mahal dalam memberi skor. Faktor-faktor tersebutlah yang menyebabkan terjadinya kesalahan dalam proses pengukuran.

Kesalahan dalam pengukuran merupakan hal yang biasa terjadi dalam pembelajaran, namun hal ini sebisa mungkin harus diminimalisir agar kesalahan pengukuran yang terjadi tidak begitu mempengaruhi keberhasilan pembelajaran. Salah satu faktor yang paling mendasar dalam kesalahan pengukuran adalah alat ukur itu sendiri. Salah satu langkah yang bisa guru lakukan untuk meminimalisir hal tersebut, guru harus bisa membuat alat ukur yang valid dan reliabel, agar alat ukur tersebut bisa menghasilkan kesalahan pengukuran yang sekecil mungkin. Pada umumnya untuk membuat alat ukur yang memiliki tingkat kesalahan pengukuran yang kecil tidak mudah, guru harus selalu mengevaluasi dan memperbaiki alat ukur yang dibuatnya dengan cara selalu melakukan analisis butir soal secara kontiniu dan mengestimasi besarnya kesalahan pengukuran dari alat ukur yang dirancang atau dibuat oleh guru.

Banyak metode yang berkembang yang bisa digunakan untuk mengestimasi besarnya kesalahan pengukuran dalam pengujian baik pada Teori Tes Klasik maupun Teori Respon Butir. Perbedaan asumsi dan teknik yang terdapat pada masing-masing tersebut menyebabkan hasil estimasi kesalahan pengukuran bervariasi dan juga menghasilkan akurasi yang berbeda. Penelitian tentang estimasi kesalahan pengukuran soal-soal UAS Matematika Kelas XII SMAN/ sederajar di Kota Selatpanjang Kabupaten Provinsi Riau menggunakan metode Thondike, Metode *compoun binomial*, Metode Analisis Varian dan Metode Teori Respon Butir

Kesalahan baku pengukuran merupakan estimasi tentang bagaimana seorang peneliti evaluasi mengharapakan kesalahan dari tes yang telah dibuat (Sukardi, 2008, p. 50). Kesalahan baku pengukuran pada umumnya dapat juga menunjukkan tingkat reliabilitas tes. Jika nilai kesalahan baku pengukuran suatu tes yang telah dibuat kecil, berarti reliabilitas tes tersebut tinggi.

Kesalahan baku pengukuran atau yang dikenal dengan *Standard Error of Measurement (SEM)* memiliki hubungan yang erat dengan kegiatan pengukuran atau hasil pengukuran dan kemampuan/pemahaman masing-masing peserta didik pada studi yang mereka pelajari.

Hal ini didukung oleh teori tes klasik yang menyatakan nilai observasi (X) merupakan hasil penjumlahan antara nilai sebenarnya (T) dan kesalahan baku dalam pengukuran (E). Sehingga dapat disimpulkan bahwa, kesalahan baku dalam pengukuran merupakan penyimpangan yang terjadi antara skor yang tampak dan skor yang sebenarnya.

Terdapat berbagai ahli yang mengemukakan pengertian tentang kesalahan baku pengukuran. Chadha (2009, p.22) mengemukakan bahwa "*standard error of measurement/ SEM is the difference between the actual score of a person on a certain job and the obtained score by him*". Senada pendapat tersebut Crocker & Algina (1986, p.110) menyatakan "*error of measurement is the discrepancy between an examinee's observed test score and his or her true score*". Berdasarkan pernyataan-pernyataan tersebut dapat disimpulkan bahwa kesalahan baku pengukuran merupakan nilai penyimpangan yang diperoleh peserta tes atas nilai tampak mereka dengan nilai yang sebenarnya. Besarnya penyimpangan tersebut sangat bervariasi antara perangkat tes yang satu dan yang lainnya. Semakin kecil kesalahan baku pengukuran, dapat dikatakan bahwa perangkat tes tersebut memiliki error yang kecil, sehingga sangat baik jika digunakan dalam melakukan pengukuran.

Berbeda dengan pendapat tersebut di atas Kubiszyn & Gary (2010, p.253) menyatakan bahwa "*The standard error of measurement of a test is the standar deviation of error score of the test*". Makna tersebut dapat diartikan bahwa kesalahan baku pengukuran pada merupakan standar deviasi dari skor error yang terdapat pada tes. Berdasarkan pernyataan tersebut dapat disimpulkan bahwa kesalahan baku pengukuran merupakan standar deviasi yang terdiri dari skor-skor error.

Kesalahan baku pengukuran (*SEM*) ada dua kategori, yaitu kesalahan pengukuran secara acak dan kesalahan pengukuran secara sistematis. Kesalahan pengukuran secara acak bisa disebabkan oleh kondisi fisik, mental yang kurang baik. Berhubungan dengan fisik, misalnya siswa ketika melakukan pengukuran atau tes sedang dalam keadaan tidak fit atau dengan kata lain, siswa tersebut sedang sakit. Berhubungan mental misalnya, ketika melakukan pengukuran siswa dihadapi permasalahan yang membuat mentalnya terganggu (dimarahi orang tua, terjadi pertengkaran, dan masalah mental lainnya). Berhubungan dengan lingkungan

misalnya, ketika pelaksanaan pengukuran berlangsung, kondisi lingkungan kurang nyaman, lokasi sekolah penuh dengan keramaian. Hal-hal tersebut bisa menyebabkan kesalahan dalam pengukuran.

Kesalahan pengukuran yang sistematis adalah kesalahan pengukuran yang disebabkan oleh alat ukur itu sendiri. Hal tersebut berkaitan dengan koefisien reliabilitas suatu alat ukur yang digunakan dalam melakukan pengukuran. Miller (2008, p.93) menyatakan bahwa “*SEM is a quantitative expression of the magnitude of error in test score based on the test reliability*”. Berdasarkan pengertian tersebut dapat diartikan bahwa kesalahan baku pengukuran adalah ekspresi kuantitatif dari banyaknya kesalahan di dalam skor tes berdasarkan pada reliabilitas. Pernyataan tersebut mensinyalir bahwa kesalahan baku dalam pengukuran sangat bergantung pada kesalahan yang disebabkan oleh alat tes itu sendiri, dalam hal ini adalah reliabilitas suatu tes. Semakin tinggi tingkat koefisien dari reliabilitas skor tes, maka semakin kecil kesalahan baku pengukurannya, sebaliknya semakin rendah tingkat koefisien reliabilitas skor tes, maka semakin besar kesalahan baku dalam pengukuran.

Kesalahan-kesalahan dalam pengukuran merupakan hal yang wajar, hal ini dikarenakan untuk membuat perangkat tes atau alat ukur yang benar-benar reliabel adalah hal yang sulit. Kubiszyn & Gary (2010, p.351) menyatakan bahwa “*The Notion of error in testing is very similar. No test measures perfectly, and many tests fail to measure as well as we would like them to*”. Pernyataan tersebut dapat diartikan bahwa, gagasan dari error atau kesalahan dalam pengujian adalah sangat serupa. Tidak ada tes yang mengukur secara sempurna dan banyak tes gagal dalam mengukur. Senada dengan pernyataan tersebut Chadha (2008, p.22) menyatakan bahwa “*making error is quite natural because nothing can be labeled as ‘complete’ or ‘perfect’ it self*”. Berdasarkan pernyataan tersebut dapat diartikan bahwa membuat kesalahan adalah hal yang sungguh alami karena tidak ada satupun manusia yang kompleks dan sempurna.

#### Metode Thorndike

Metode Thorndike memiliki asumsi bahwa tidak ada korelasi antara  $E_1$  dan  $E_2$ . Asumsi ini dipopulerkan oleh Feldt, Steffen, dan Gupta (1985, p.352) yang menyatakan

bahwa “*the correlation between  $E_1$  dan  $E_2$  is assumed equal to zero*. Hal tersebut yang mendasari munculnya Rumus Thorndike dalam menentukan kesalahan dalam pengukuran. Persamaan yang mendasari munculnya estimasi kesalahan dalam pengukuran adalah sebagai berikut:

$$\sigma_E^2 = \sigma_{E1}^2 + \sigma_{E2}^2 \dots \dots \dots (22)$$

$$\sigma_E = (\sigma_{E1}^2 + \sigma_{E2}^2)^{1/2} \dots \dots \dots (23)$$

Besar nilai estimasi kesalahan dalam pengukuran dengan Metode Thorndike dapat dicari dengan menghitung nilai dari  $\sigma_{E1}^2$  dan  $\sigma_{E2}^2$ . Metode Thorndike memiliki hubungan erat dengan Metode Teori Tes Klasik sehingga penentuan besarnya nilai  $\sigma_{E1}^2$  dan  $\sigma_{E2}^2$  dapat dilakukan dengan berpedoman pada rumus penentuan besarnya SEM berdasarkan Teori Tes Klasik seperti pada rumus berikut ini.

$$SEM = SD \sqrt{1 - r_{xx}} \dots \dots \dots (24)$$

$$\sigma_{E1} = \sigma_{E2} \sqrt{1 - r_{xx}} \dots \dots \dots (25)$$

$$\sigma_{E2} = \sigma_{E2} \sqrt{1 - r_{xx}} \dots \dots \dots (26)$$

Rumus tersebut merupakan rumus awal dalam menentukan estimasi kesalahan baku dalam pengukuran dengan menggunakan Metode Thondike. Selanjutnya baru disubsitisi ke dalam Rumus Metode Thorndike.

#### Metode Compound Binomial

Metode *compound binomial* merupakan metode yang pertama kali digagas oleh Lord. Feldt, Steffan, & Gupta (1985, p.354) mereka mengemukakan bahwa *mathing forms during test construction is essentially a process of selecting stratified samples of item rather than completely random samples from the populations of item*. Pengertian tersebut dapat diartikan bahwa sebuah perangkat tes berdasarkan Metode *Compound Binomial* terdiri dari kelompok-kelompok butir soal yang memiliki karakteristik berstrata. Senada pendapat tersebut Mardapi (1999, p.26) mengemukakan bahwa suatu perangkat tes pada dasarnya merupakan suatu proses pemilihan sampel bersrta dari kumpulan butir soal. Berdasarkan pendapat kedua ahli tersebut dapat disimpulkan bahwa langkah yang membedakan Metode *Compound Binomial* dibandingkan dengan metode yang lain adalah terletak pada pemilihan butir soal menjadi sebuah perangkat soal. Perakitan sebuah soal merupakan sebuah proses penyeleksian butir-butir soal dari kumpulan butir soal yang

memiliki karakteristik tertentu dan berasal dari strata/tingkatan yang berbeda. Strata atau tingkatan dalam penyeleksian tersebut dilihat dari tingkat kesukaran suatu soal (*Item Difficulty*).

Estimasi kesalahan pengukuran dari masing-masing strata harus dihitung terlebih dahulu. Akar dari penjumlahan varians kesalahan baku pengukuran masing-masing strata merupakan nilai kesalahan baku pengukuran perangkat tes tersebut. berdasarkan teori yang dikemukakan Lord di atas, perakitan perangkat tes bukan pengambilan secara acak butir soal pada sebuah populasi butir soal. Rumus yang dikemukakan oleh Lord untuk menghitung besarnya kesalahan pengukuran pada sebuah perangkat tes adalah sebagai berikut.

$$S_{E(i)} = \left[ \sum_{h=1}^c \frac{X_{ih}(k_h - X_{ih})}{k_h - 1} \right]^{1/2} \dots\dots\dots(27)$$

Keterangan rumus:

- $S_{E(i)}$  : Kesalahan pengukuran bagi peserta tes  $i$
- $X_{ih}$  : Skor yang diperoleh peserta tes ke  $i$  pada rumpun atau kelompok dari butir tes pada keategori  $h$  pada sebuah spesifikasi perangkat tes .
- $k_h$  : Jumlah butir soal pada kategori  $h$

**Metode Analisis Varians**

Metode Analisis Varians (ANAVA) merupakan suatu metode yang praktis. Hal ini dikarenakan perangkat tes tidak perlu dibagi menjadi lebih dari satu bagian dan memperhatikan besarnya nilai rerata dan baian. Metode ini ditemukan oleh Hoyt. Menurut Feld, Steffan, Gupta (1985, p.354) Asumsi dasar yang digunakan oleh Hoyt dalam mengestimasi kesalahan pengukuran menggunakan Metode Analisis Varian adalah bahwa *the examinees by items score matrix for a test may be analyzed to obtain mean squares for examinees (MS<sub>S</sub>) items (MS<sub>I</sub>) dan interaction (MS<sub>SxI</sub>) and the reliability of the test may be estimated from these mean squares.* Pernyataan tersebut dapat diartikan bahwa sekor/nilai butir soal dapat digunakan untuk menghitung rerata kuadrat peserta tes, rerata kuadrat item soal, serta rerata kuadrat interaksi antara peserta tes dan butir soal. Besarnya koefisien reliabilitas dan kesalahan baku pengukuran juga dapat dihitung berdasarkan hasil perhitungan rerata kuadrat tersebut.

Azwar (2010, p.92) mengemukakan bahwa konsep dari teknik ini adalah meman-

dang distribusi jawaban peserta tes sebagai suatu desain eksperimen faktorial dua jalan yang dikenal pula dengan *item by subject design*. Setiap butir soal dianggap sebagai treatment atau perlakuan. Berdasarkan metode ini, variasi besarnya kesalahan pengukuran pada sebuah perangkat tes yang terdiri dari  $k$  item dapat diperkirakan dengan menggunakan nilai  $MS_{SxI}$ . Mempermudah pemahaman pernyataan tersebut, rumus yang dikemukakan oleh Hoyt dalam penentuan kesalahan baku pengukuran sebagai berikut.

$$S_E = [k(MS_{SxI})]^{1/2} \dots\dots\dots(28)$$

Keterangan rumus:

- $S_E$  : Kesalahan Pengukuran
- $K$  : Jumlah Butir Soal Sebuah perangkat tes
- $MS_{SxI}$ : *Means Square* interaksi antara peserta tes dan butir soal

Azwar (2010, p.93) mengemukakan rumus sederhana dalam menentukan besarnya kesalahan pengukuran sebagai berikut.

$$MS_{SxI} = \frac{\left( i - \frac{\sum X^2}{k} - \frac{\sum Y^2}{n} \right) + \frac{\sum i^2}{nk}}{(n-1)(k-1)} \dots\dots\dots(29)$$

Keterangan rumus:

- $MS_{SxI}$ : *Means Square* interaksi antara peserta tes dan butir soal
- $X$  : Nilai yang diperoleh masing-masing peserta tes pada keseluruhan butir soal
- $Y$  : Nilai keseluruhan peserta tes pada satu butir soal

**Metode Teori Respon Butir**

Metode estimasi kesalahan baku pengukuran Teori Respon Butir merupakan metode yang menuntut ukuran sampel yang lebih besar agar hasil yang diperoleh akurat. Nilai fungsi informasi sangat dibutuhkan dalam mencari kesalahan pengukuran dalam Teori Respon Butir (Hamleton, Swaminathan, & Rogers, 1991, p.37). Secara sistematis dapat ditulis,

$$SEM(\theta) = \frac{1}{\sqrt{I(\theta)}} \dots\dots\dots(30)$$

Keterangan rumus:

- $SE(\theta)$ : Kesalahan baku dalam pengukuran
- $I(\theta)$  : *Harga* fungsi informasi tes terhadap parameter tingkat kemampuan peserta tes.
- $\theta$  : *Tingkat* kemampuan peserta tes.

Kesalahan baku dalam pengukuran pada teori tes klasik berpatokan kepada tingkat reliabilitas suatu soal. Semakin tinggi tingkat reliabilitas suatu soal maka semakin rendah kesalahan pengukurannya. Sedangkan dalam teori respon butir kesalahan baku dalam pengukuran berpatokan pada fungsi informasi suatu tes, semakin besar fungsi informasi suatu tes maka semakin kecil kesalahan dalam pengukuran. sehingga dapat disimpulkan bahwa besar kecilnya nilai fungsi informasi suatu tes sangat mempengaruhi kesalahan baku dalam pengukuran. semakin besar fungsi informasi maka tes tersebut cermat dalam mengukur kemampuan peserta tes. kecermatan hasil ukur yang ditunjukkan dengan lebar sempitnya interval kepercayaan bagi kemampuan peserta yang sebenarnya (T).

### Metode Penelitian

Penelitian ini memiliki tujuan untuk mengestimasi besarnya kesalahan baku pengukuran perangkat tes buatan guru pada UAS mata pelajaran Matematika kelas XII pada jurusan IPA dan IPS tahun ajaran 2013/2014 yang digunakan oleh sekolah-sekolah SMA/ sederajat di Kota Selatpanjang Kabupaten Kepulauan Meranti Provinsi Riau. Berdasarkan tujuan tersebut maka jenis penelitian ini merupakan penelitian deskriptif kuantitatif. Penelitian ini dikategorikan sebagai penelitian deskriptif kuantitatif karena penelitian ini mendeskripsikan atau menggambarkan data yang telah dikumpulkan dengan tujuan untuk mendeskripsikan besarnya kesalahan pengukuran yang terjadi dalam proses pengujian dengan menggunakan Metode Thorndike, Metode *compound binomial*, Metode Analisis Varians dan Metode Teori Respons Butir. Tempat penelitian dalam penelitian *Estimasi Kesalahan Pengukuran Pada Soal-Soal UAS kelas XII mata pelajaran Matematika* adalah di Kota Selatpanjang di Kabupaten Kepulauan Meranti Provinsi Riau yang dilaksanakan mulai tanggal 20 Januari 2014.

Objek dalam penelitian ini adalah seluruh lembar jawaban siswa Ujian Akhir Semester kelas XII mata pelajaran Matematika SMA pada jurusan IPA dan IPS yang dirancang oleh guru tahun ajaran 2013/2014. LJK yang digunakan dalam penelitian ini diperoleh dari 4 SMA yang berada di kota Selatpanjang Kabupaten Kepulauan Meranti yang memiliki perangkat tes yang berbeda pada setiap sekolah.

Lembar jawaban siswa tersebut dianalisis untuk melihat besarnya kesalahan pengukuran yang terjadi perangkat tes yang dibuat oleh masing-masing.

Teknik pengumpulan data dalam penelitian estimasi kesalahan pengukuran pada soal-soal UAS mata pelajaran matematika SMA Kelas XII di Kota Selatpanjang Kabupaten Kepulauan Meranti Provinsi Riau adalah dengan menggunakan dokumentasi. Peneliti langsung mendatangi setiap sekolah SMA di kota Selatpanjang Kabupaten Kepulauan meranti dan menemui kepala sekolah meminta izin agar bisa mengambil data di sekolah yang bersangkutan, kemudian peneliti menemui guru mata pelajaran matematika yang mengetahui banyak tentang soal dan lembar jawaban siswa yang telah diujikan pada tahun ajaran 2013/2014. Data yang diambil adalah lembar jawaban siswa dan soal pada Ujian Akhir Semester tahun ajaran 2013/2014.

Data diestimasi dengan menggunakan metode Thondike, Metode *compoun Binomial*, Metode Analisis Varian, Metode Teori Respon Butir untuk melihat besarnya kesalahan pengukuran yang terjadi pada masing-masing perangkat soal yang dirancang oleh guru-guru SMAN/sederajat di Kota Selatpanjang Kabupaten Kepulauan Meranti Provinsi Riau. Pelaksanaan analisis dilakukan dengan menggunakan bantuan komputer *Microsoft Excell*.

### Hasil Penelitian dan Pembahasan

Berdasarkan hasil analisis yang telah dilakukan dengan menggunakan Metode Thondike, Metode *Compound Binomial*, Metode Analisis Varian, dan Metode Teori Respon Butir diketahui bahwa estimasi kesalahan pengukuran yang paling kecil terdapat pada Metode *CompoundBinomial* yaitu pada perangkat tes rancang guru MAN Jurusan IPA. Rangkuman estimasi kesalahan pengukuran selengkapnya dapat dilihat pada Tabel 1.

Estimasi kesalahan dalam pengukuran dengan Metode Thondike dapat dicari dengan menghitung nilai dari  $\sigma_{E1}^2$  dan  $\sigma_{E2}^2$ . Metode Thondike memiliki hubungan erat dengan Metode Teori Tes Klasik sehingga penentuan besarnya nilai  $\sigma_{E1}^2$  dan  $\sigma_{E2}^2$  dapat dilakukan dengan berpedoman pada rumus penentuan besarnya SEM berdasarkan Teori Tes Klasik. Penentuan varian error setiap belahan erat kaitannya dengan nilai reliabilitas perangkat tes

tersebut, oleh karena itu nilai reliabilitas perlu dicari sesuai dengan hasil pembelahan dari masing-masing perangkat tes.

Tabel 1. Rangkuman Hasil Estimasi Kesalahan Baku

Metode	SMAN1		MAN 1	
	IPA	IPS	IPA	IPS
Thondike	1,935	2,417	2,921	2,686
<i>Compoun binomial</i>	1,005	1,079	0,917	0,983
ANOVA	2,170	2,108	2,374	2,672
1-PL	0,418 s.d 0,850	0,413 s.d 0,902	0,332 s.d 0,644	0,330 s.d 0,700
IIRT 2-PL	0,682 s.d 1,070	0,551 s.d 1,516	0,513 s.d 0,673	0,540 s.d 1,929
3-PL	0,548 s.d 2,290	0,360 s.d 2,232	0,242 s.d 0,932	0,240 s.d 1,025

Tabel 2. Estimasi kesalahan baku pengukuran

Metode	SMAN2		SMAN3	
	IPA	IPS	IPA	IPS
Thondike	2,421	2,398	2,974	2,335
<i>Compoun binomial</i>	1,066	1,039	1,262	1,104
ANOVA	2,492	2,525	2,193	2,533
1-PL	0,357 s.d 0,776	0,350 s.d 0,797	0,349 s.d 2,272	0,330 s.d 0,728
IIRT 2-PL	0,631 s.d 1,043	0,605 s.d 0,950	0,579 s.d 0,954	0,570 s.d 0,936
3-PL	0,528 s.d 1,660	0,432 s.d 1,845	0,318 s.d 0,783	0,573 s.d 1,091

Estimasi kesalahan baku pengukuran Metode *compoun binomial* tidak jauh berbeda dengan Metode Thondike. Perbedaan Metode *compoun binomial* dengan Metode Thondike hanya terletak pada proses pembelahan butir soal perangkat tes yang diujikan. Pembelahan butir soal pada Metode *compoun binomial* menekankan kepada kelompok butir tes yang berstrata. Hasil pembelahan dari butir soal yang berstrata tersebut dicari kesalahan pengukuran individu pada masing-masing strata. Estimasi kesalahan kelompok pesertea tes dilakukan dengan mengkuadratkan jumlah varian kesalahan an individu pada masing-masing strata.

Estimasi kesalahan baku pengukuran berdasarkan Metode analisis varian dilakukan dengan menghitung interaksi peserta tes dengan butir soal. Metode ini berasumsi bahwa setiap jawaban yang diberikan oleh peserta tes pada

setiap butir soal merupakan suatu desain eksperimen faktorial dua jalan. Estimasi kesalahan pengukuran tersebut dapat dilakukan dengan menghitung *means square* interaksi antara peserta tes dengan butir soal yang dijawabnya.

Estimasi kesalahan baku pengukuran perangkat tes berdasarkan Metode Teori Respon Butir dilakukan dengan menghitung fungsi informasi pada perangkat tes. perhitungan fungsi informasi dilakukan karena fungsi informasi sangat erat kaitannya dengan kesalahan baku pengukuran. fungsi informasi berbanding terbalik dengan kesalahan baku pengukuran, semakin besar nilai fungsi informasi maka semakin kecil kesalahan pengukuran pada suatu perangkat tes. Estimasi kesalahan baku pengukuran perangkat tes SMAN/ sederajat di Kota Selatpanjang Kabupaten Kepulauan Meranti Provinsi Riau berdasarkan Teori Respon Butir dilakukan dengan menggunakan 3 Model logistik, yaitu model logistik 1-Parameter, model logistik 2-Parameter, dan model logistik 3-Parameter.

Estimasi kesalahan yang diperoleh dari berbagai metode dalam mengestimasi kesalahan pengukuran dapat digunakan untuk mencari skor murni yang diperoleh siswa dari proses pengujian. Lebar sempitnya interval skor murni tergantung pada besar kecilnya nilai kesalahan pengukuran. Kesalahan baku pengukuran membantu membantu menjelaskan batas kesalahan yang terjadi pada setiap tes dan secara khusus berguna untuk mengertimasi skor peserta tes yang sebenarnya. Kesalahan baku pengukuran membantu memperlihatkan jarak antara skor murni dan skor tampak. Semakin besar kesalahan baku pengukuran semakin lebar interval skor murni, yang maksudnya alat ukur yang digunakan tidak cermat. Sebaliknya, semakin kecil kesalahan baku pengukuran semakin kecil pula interval skor murni dan hasil pengukuran semakin cermat.

## Simpulan dan Saran

Berdasarkan hasil penelitian dan pembahasan yang telah disajikan dalam BAB sebelumnya, terdapat lima kesimpulan yang dihasilkan dalam penelitian ini. Kelima kesimpulan tersebut adalah sebagai berikut.

1. Berdasarkan Metode Thorndike besarnya estimasi kesalahan baku pengukuran perangkat tes UAS mata pelajaran Matematika SMAN1, MAN, SMAN 2, dan

- SMAN 3 jurusan IPA dan IPS secara berturut adalah sebesar, 1,953, 2,417, 2,921, 2,686, 2,421, 2,398, 2,974, 2,335, kesalahan baku pengukuran terkecil terjadi pada perangkat tes SMAN1 jurusan IPA, sedangkan kesalahan baku pengukuran terbesar terjadi pada perangkat tes SMAN 3 jurusan IPA.
2. Berdasarkan Metode *Compound Binomial* estimasi kesalahan baku pengukuran perangkat tes UAS mata pelajaran Matematika SMAN1, MAN, SMAN2, SMAN 3 jurusan IPA dan IPS secara berturut-turut adalah sebesar, 1,005, 1,079, 0,917, 0,983, 1,066, 1,039, 1,262, 1,104, kesalahan baku pengukuran terkecil berdasarkan Metode *Compound Binomial* adalah perangkat tes MAN Jurusan IPS, sedangkan kesalahan baku pengukuran yang terbesar adalah perangkat tes SMAN 3 jurusan IPA.
  3. Berdasarkan Metode Analisis Varians estimasi kesalahan baku pengukuran perangkat tes UAS mata pelajaran Matematika SMAN 1, MAN, SMAN 2, SMAN 3 Jurusan IPA dan IPS secara berturut adalah sebesar 2,170, 2,108, 2,374, 2,672, 2,492, 2,525, 2,193, 2,533, kesalahan baku pengukuran terkecil berdasarkan Metode Analisis Varians adalah perangkat tes SMAN 3 Jurusan IPS, sedangkan kesalahan baku pengukuran terbesar adalah perangkat tes SMAN 1 Jurusan IPS.
  4. Berdasarkan Metode Teori Respons Butir Besarnya estimasi kesalahan baku pengukuran individu terkecil sampai terbesar dengan menggunakan tiga model logistik sebagai berikut.
    - a. Kesalahan baku pengukuran individu Model Logistik 1-Parameter terkecil terdapat pada perangkat tes MAN Jurusan IPA, yaitu pada  $\theta=0$  dengan nilai sebesar 0,322 dan nilai kesalahan individu terbesar terdapat pada perangkat tes SMAN 2 Jurusan IPS yaitu pada  $\theta=-3$  dengan nilai sebesar 0,797
    - b. Kesalahan baku pengukuran individu Model Logistik 2-Parameter terkecil terdapat pada perangkat tes MAN Jurusan IPA, yaitu pada  $\theta=-0,4$  dengan nilai sebesar 0,513 dan nilai kesalahan individu terbesar terdapat pada perangkat tes SMAN 1 Jurusan IPS yaitu pada  $\theta=-3$  dengan nilai sebesar 1,516
    - c. Kesalahan baku pengukuran individu Model Logistik 3-Parameter terkecil

terdapat pada perangkat tes MAN Jurusan IPS, yaitu pada  $\theta=3$  dengan nilai sebesar 0,322 dan nilai kesalahan individu terbesar terdapat pada perangkat tes SMAN 1 Jurusan IPS yaitu pada  $\theta=-3$  dengan nilai sebesar 2,232

5. Metode yang paling akurat dalam mengestimasi kesalahan baku pengukuran soal-soal UAS Matematika kelas XII SMA di Kota Selatpanjang Kabupaten Kepulauan Meranti Provinsi Riau adalah Metode *Compound Binomial*.

Berdasarkan hasil penelitian yang telah peneliti lakukan, terdapat beberapa saran yang ingin peneliti ungkapkan sebagai berikut.

1. Secara keseluruhan soal-soal yang dirancang oleh guru SMAN/ sederajat di Kota Selatpanjang dikategorikan kurang baik. Guru diharapkan bisa membuat tes yang valid dan reliable, menganalisis secara kuantitatif soal secara kontiniu, dan mengikuti pelatihan-pelatihan peningkatan kompetensi pedagogik guru, agar guru mampu membuat soal yang valid, reliable, dan memiliki kesalahan baku pengukuran sekecil mungkin,
2. Guru atau pendidik hendaknya memiliki wawasan atau pengetahuan tentang mengestimasi kesalahan baku pengukuran agar bisa mencari nilai murni/*true score* peserta didiknya.
3. Guru atau pendidik hendaknya melakukan analisis kesalahan baku pengukuran dari setiap perangkat tes yang telah dirancangnya secara kontiniu.
4. Guru atau pendidik yang ingin mengetahui kesalahan baku pengukuran pada perangkat tes yang diujikan kepada pesertanya bisa mengestimasi dengan menggunakan Metode Thorndike, Metode *Compound Binomial*, Metode Analisis Varians dan Metode Teori Respons Butir.
5. Pengetahuan guru tentang estimasi kesalahan baku pengukuran masih sangat kurang sekali, oleh karena itu penelitian pada responden yang berbeda dalam provinsi atau kabupaten yang berbeda perlu dilakukan.

#### Daftar Pustaka

- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey: Wadsworth.

- Campbell, C., & Levin, B. (2009). Using data to support educational improvement. *Journal Educational Assessment, evaluation and accountability*: Springer Sciences + Bussiness Media, LLC.
- Chadha, N. K. (2009). *Applied psykometry*. New Delhi: Sage Publications Inc.
- Djemari Mardapi. (2008). *Teknik Penyusunan intrumen tes & non tes*. Yogyakarta: Mitra Cendikia
- Djemari Mardapi. (2012). *Pengukuran Penilaian & Evaluasi Pendidikan*. Yogyakarta: Nuha Litera
- Farida Yusuf Tayibnapi. (2000). *Evaluasi Program dan Intrumen Evaluasi*. Jakarta: Rineka Cipta
- Feldt, L. S. (1995). Estimation of reliability of differences under revised reliabilities of component scores. *Journal of educational measurement*: Susan L. Wantland.
- Gardner, J. (2012). *Assessment and learning. 2<sup>th</sup> edition*. California: Sage Publications Ltd
- Gronlund, N. E. (1977). *Constructing Achievement Test*. London: Printice-Hall.
- Gronlund, N. E., & Linn, R. L. (1990). *Measurement and Evaluation in Teaching. 6<sup>th</sup> edition*. New York: Collier Macmillan Publishers.
- Hamzah B. Uno dan Satria Koni. *Assessment Pembelajaran*. Jakarta: Bumi Aksara.
- Harun Rasyid & Mansur. (2008). *Penilaian hasil belajar*. Bandung: Wacana Prima.
- Hambleton, R. K., Swaminathan H., & Rogers H. Jane. (1991). *Fundamentals of Item Response Theory*. Newbury Park: Sage Publications.
- Heritage, M. (2010). *Formative assessment. Making it happen in the classroom*. United States of America: A Sage Company
- Kubiszyn, T., & Borich, G. (2010). *Educational Testing & Measurement Classroom Application and Prakte: United State of America*: RRD Crawfordsvelle.
- Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement And Evaluation In Education*. United States Of America: Thomson Learning Academic Resource Center.
- Miller, P. W. (2008). *Measurement and teaching*. Muster: Partric W. Miller & Association.
- Miller, D., Linn, R. L., & Gronlund, N. E. (2009) *Measurement and assessment in teaching*. United State of America: Pearson.
- Phye. (1997). *Handbook of Classroom Asseessment, learning, adjustment, and achievement*. New York: Academi press.
- Reynolds, C. R., Livingston, R. B., & Wilson, V. (2010). *Measurement and assessment in education*. United States of America: Pearson.
- Sukardi. 2008, *Evaluasi Pendidikan Prinsip dan Operasionalnya*. Jakarta: Bumi Aksara
- Suharsimi Arikunto. (2012). *Dasar-dasar evaluasi pendidikan*. Yogyakarta: Bumi Aksara
- Wiliam, D. (2009). formatif assessment getting the focus right. *Journal of education Testing service*: Lawrence Erlbaum Associates.
- Wirawan. (2011). *Evaluasi teori, model, standar, aplikasi, dan profesi*. Jakarta: Rajagrafindo Persada.
- Worthen, B. R., & Sanders, J. R. (1973). *Educational Evaluation Theory and Practice*.: University Of Colorado. Charles A. Jones Publishing Company.
- Wright, R. J. (2008) *Eduacational Assessment*. United States of America: Sage Publications.