

ESTIMASI KESALAHAN PENGUKURAN BUTIR SOAL UJIAN AKHIR SEMESTER MATEMATIKA

Devi Dwi Kurniawan

Program Studi Manajemen Fakultas Ekonomika dan Bisnis Universitas Selamat Sri

devidwikurniawan@gmail.com

ABSTRAK

Studi ini bertujuan (1) untuk mengestimasi kesalahan pengukuran butir soal ujian akhir semester (UAS) Matematika SMA Pondok Modern Selamat, (2) untuk melihat akurasi beberapa metode estimasi kesalahan pengukuran dan (3) untuk memberikan pedoman bagi para guru atau pendidik yang menyelenggarakan tes dalam mengestimasi kesalahan pengukuran.

Analisis data yang didasarkan pada respon peserta didik terhadap perangkat butir soal ujian akhir semester Matematika SMA Pondok Modern Selamat. Sumber data berupa lembar jawaban dari 120 siswa SMA Pondok Modern Selamat Kelas XII. Analisis butir soal dilakukan dengan menggunakan program Itean untuk melihat karakteristik soal. Selanjutnya dilakukan analisis kesalahan pengukuran dengan menggunakan metode Thorndike, Metode Compound Binomial, Metode Compuond Binomial Modifikasi Keats, Metode Compound Binomial Modifikasi Feldt, dan Metode ANAVA.

Hasil analisis estimasi kesalahan pengkuran butir soal ujian akhir semester matematika SMA Pondok Modern Selamat dengan menggunakan metode Thorndike, Metode Compound Binomial, Metode Compuond Binomial Modifikasi Keats, Metode Compound Binomial Modifikasi Feldt, dan Metode ANAVA secara berturut-turut adalah 2,370; 0,752; 1,207; dan 2,21.

Kata kunci: Estimasi Kesalahan Pengukuran, Soal Tes Matematika

ABSTRACT

The objectives of the paper are using (1) to estimate the standard error of measurement (SEM) of Mathematics final examination result in Senior High School Pondok Modern Selamat, (2) to investigate the accuracy of the various methods in estimating SEM, and (3) to give the orientation for teachers who make a test in estimating SEM.

The data analysis was based on students' of the mathematics test items of final examination results in Senior High School Pondok Modern Selamat. The data source was in a form of computerised answer sheet of 120 of the third grade students' in Senior High School Pondok Modern Selamat. The analysis of the test items use the MicroCat Itean program to show the characteristic of the test items. The estimation of SEM analysis was done by using Thorndike Method, Compound Binomial Method, Compound Binomial Method Modification Keats, and ANAVA Method.

The results of the SEM analysis of the mathematics test items of final examination in Senior High School Pondok Modern Selamat used Thorndike Method, Compound Binomial Method, Compound Binomial Method Modification Keats, and ANAVA Method insequence are 2,370; 0,752; 1,207; and 2,21.

Key words: Standard Error of Measurement, Mathematics Test Items

PENDAHULUAN

Pengukuran merupakan suatu langkah yang harus diambil untuk melaksanakan evaluasi. Ketika akan dilakukan sebuah proses evaluasi, maka hendaknya didahului dengan sebuah proses pengukuran yang cermat. Hal ini berarti pengukuran merupakan sebuah cara yang harus dilakukan terlebih dahulu sebelum dilakukannya evaluasi. Baik buruknya hasil evaluasi tergantung dari hasil pengukuran. Hasil

pengukuran yang kurang cermat akan menghasilkan evaluasi yang kurang cermat, begitupun sebaliknya, hasil pengukuran yang baik akan menghasilkan evaluasi yang baik pula.

Pengukuran pendidikan merupakan kegiatan melakukan kuantifikasi gejala atau objek. Gejala atau objek tersebut bisa berupa motivasi, prestasi, percaya diri atau prestasi yang semuanya dinyatakan dalam bentuk angka (Mardapi, 2012:1). Keberhasilan dalam proses pembelajaran dapat dilihat melalui proses penilaian. Dalam usaha menilai prestasi belajar dari para siswa, biasanya guru melakukan pengukuran dalam bentuk ujian-ujian baik tertulis, praktek maupun lisan. Dari hasil ujian-ujian itu guru memberikan skor, yang lazimnya skor tersebut dalam bentuk angka-angka. Namun kemudian timbul pertanyaan, apakah skor yang didapatkan dari hasil ujian tersebut merupakan skor siswa yang sebenarnya?, Skor sebenarnya yang dimaksud adalah skor yang sebenarnya mencerminkan prestasi belajar dari siswa yang bersangkutan. Secara umum, kita menyadari adanya kemungkinan terdapat perbedaan skor hasil pengukuran dengan skor yang sebenarnya.

Wright (2008:130) *True score = observed score $\pm$ measurement error*. Dari persamaan tersebut ada dua kemungkinan yang akan terjadi. Pertama, mungkin skor hasil pengukuran lebih kecil dari skor sebenarnya. Kedua, bisa jadi skor hasil pengukuran lebih besar dari skor sebenarnya. Bila salah satu dari kedua kemungkinan tersebut terjadi, berarti pada proses pengukuran telah terjadi kesalahan.

Untuk menghilangkan semua sumber penyebab terjadinya kesalahan pengukuran sangat sulit, tetapi dapat diusahakan agar kesalahan pengukuran dapat diminimalkan, sehingga perolehan skor dapat mencerminkan kemampuan peserta tes yang sebenarnya. Salah satu faktor yang paling mendasar dalam kesalahan pengukuran adalah alat ukur itu sendiri. 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.

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, 2013:2). Senada dengan pendapat tersebut Mardapi (2008:67) menyatakan untuk menghasilkan informasi yang akurat, maka instrumen dalam pengukuran harus handal, sehingga instrument tersebut mampu menghasilkan kesalahan baku pengukuran yang sekecil mungkin.

Banyak metode yang berkembang yang bisa digunakan untuk mengestimasi besarnya kesalahan baku pengukuran. Pada prinsipnya metode ini dapat dikategorikan menjadi dua, yaitu yang dikembangkan berdasarkan Teori Tes Klasik dan Teori Respons Butir. Masing-masing model memiliki asumsi dan formula yang berbeda-beda. Pemilihan model tersebut tergantung dari asumsi yang diterapkan pada seperangkat tes yang akan diestimasi kesalahan pengukurannya.

Evaluasi pendidikan pada jenjang SMA dilakukan setiap 6 bulan sekali, yang biasa disebut Ulangan Akhir Semester (UAS). Pelaksanaan evaluasi semacam ini didasarkan pada sistem yang diterapkan yang menganut sistem semester. Bagi guru SMA kelas XII, Ulangan Akhir Semester I diharapkan dapat untuk melihat kesiapan anak dalam menghadapi Ujian Nasional (UN), karena materi kelas XII umumnya merupakan 50% dari total materi soal UN di SMA.

Agar hasil UAS dapat menggambarkan hasil yang sesungguhnya, maka perangkat soal yang digunakan harus memenuhi segala persyaratan bagi sebuah alat ukur yang baik dan teruji dalam berbagai aspek. Permasalahan yang muncul sekarang apakah soal-soal UAS benar-benar merupakan alat ukur yang baik yang mampu mencerminkan kemampuan pesertanya. Banyak faktor yang mempengaruhi keadaan

tersebut, diantaranya adalah kesalahan pengukuran. Oleh karena itu, setiap hasil pengukuran harus disertai dengan informasi tentang besarnya kesalahan pengukuran.

Meskipun pelaksanaan UAS telah dilakukan beberapa kali, namun hasil yang diperoleh belum memuaskan. Hal ini ditunjukkan dengan perolehan nilai UAS yang relative rendah, khususnya pada pelajaran Matematika. Rendahnya nilai UAS kemungkinan karena siswa kesulitan mengerjakan soal UAS. Hal ini mengindikasikan tingkat kesukaran soal UAS tidak sesuai dengan kemampuan peserta, daya beda dan pengecoh dalam soal UAS belum mampu berfungsi dengan baik. Besarnya indeks kesukaran dan daya beda butir tes mempengaruhi indeks reliabilitas tes. Tes yang mudah atau sulit cenderung memiliki nilai reliabilitas yang kecil. Kondisi ini menyebabkan terjadinya kesalahan pengukuran yang cukup tinggi. Berdasarkan hal tersebut perlu diadakan sebuah penelitian untuk melihat besarnya kesalahan baku dalam pengukuran.

Secara umum masih banyak guru yang belum melakukan analisis kesalahan baku pengukuran dalam proses pengukuran yang mereka lakukan. Hal tersebut berdasarkan kegiatan prasurvei yang dilakukan oleh peneliti di SMA Pondok Modern Selamat. Sebagian besar guru mengaku tidak mengetahui adanya kesalahan baku pengukuran, padahal analisis kesalahan baku pengukuran merupakan hal yang sangat penting karena untuk mengetahui apakah soal yang telah dibuat sudah memenuhi karakteristik baik sehingga hasil pengukuran bisa menggambarkan kemampuan siswa yang sebenarnya. Jadi, dengan adanya penelitian ini diharapkan dapat memberi kontribusi besar bagi perkembangan dan kemajuan pendidikan Khususnya dalam penilaian pembelajaran matematika Sekolah Menengah Atas (SMA).

Untuk melihat apa saja permasalahan yang akan dijawab dalam penelitian Estimasi Kesalahan Pengukuran soal-soal ujian akhir semester Matematika kelas XII SMA Pondok Modern Selamat di Kabupaten Batang Provinsi Jawa Tengah, maka perlu dirumuskan masalah-masalah yaitu, bagaimana karakteristik soal-soal UAS kelas XII mata pelajaran Matematika di SMA Pondok Modern Selamat tahun ajaran 2018/2019?; Berapa besar estimasi kesalahan baku pengukuran soal-soal UAS kelas XII mata pelajaran Matematika di SMA Pondok Modern Selamat tahun ajaran 2018/2019?; dan Bagaimanakah akurasi metode-metode estimasi kesalahan baku pengukuran dari tiga metode (*Metode Compound Binomial*, *Metode Analisis Varians* dan *Metode Teori Respons Butir*) dalam mengestimasi besarnya kesalahan baku pengukuran soal-soal UAS kelas XII mata pelajaran Matematika di SMA Pondok Modern Selamat tahun ajaran 2018/2019?

Berdasarkan rumusan masalah yang telah dibuat, maka tujuan dari penelitian ini adalah mengetahui karakteristik soal-soal UAS kelas XII mata pelajaran Matematika di SMA Pondok Modern Selamat tahun ajaran 2018/2019, mengetahui besarnya estimasi kesalahan baku pengukurun soal-soal UAS kelas XII mata pelajaran Matematika di SMA Pondok Modern Selamat tahun ajaran 2018/2019, dan mengetahui metode manakah yang paling akurat dalam mengestimasi besarnya kesalahan baku pengukuran soal-soal UAS mata pelajaran Matematika di SMA Pondok Modern Selamat Tahun Ajaran 2018/2019.

METODE PELAKSANAAN

Penelitian ini memiliki tujuan untuk mengestimasi besarnya kesalahan baku pengukuran perangkat tes buatan guru pada UAS mata pelajaran Matematika SMA Pondok Modern Selamat tahun ajaran 2018/2019 di Kabupaten Batang Provinsi Jawa Tengah. Berdasarkan tujuan tersebut maka jenis penelitian ini merupakan penelitian *Ex post facto*.

Penelitian ini dikategorikan sebagai penelitian *Ex post facto* karena penelitian ini mendeskripsikan atau menggambarkan peristiwa yang sudah terjadi, berdasarkan data yang telah dikumpulkan dengan tujuan untuk mengetahui besarnya kesalahan baku pengukuran yang terjadi dalam proses pengujian dengan menggunakan Metode Thorndike, Metode *Compound Binomial*, Metode *compound Binomial* dengan Modifikasi Keats dan Metode Analisis Varian.

HASIL DAN PEMBAHASAN

Pengukuran hasil belajar merupakan kegiatan yang penting dilakukan dalam proses pendidikan. Dalam kegiatan pembelajaran, tes hasil belajar digunakan untuk melihat keberhasilan guru dalam melaksanakan pembelajaran dan untuk melihat ketercapaian tujuan pembelajaran. Oleh karena itu, alat ukur yang digunakan dalam melakukan tes hasil belajar harus memenuhi kriteria yang baik sesuai dengan standar yang ada. Alat ukur yang baik dapat dilihat dari validitas, reliabilitas, tingkat kesukaran butir soal, daya beda, efektivitas distraktor, dan kesalahan baku pengukuran soal.

Nitko dan Brookhart (2007:38) *Validity is the soundness of your interpretations and uses students' assessment result.* Miller (2008:83) *Validity is the extent to which a test measures what it was intended to measure.* Mardapi (2012:37) validitas merupakan dukungan bukti dan teori terhadap penafsiran skor tes sesuai dengan tujuan penggunaan tes. Jadi, validitas adalah ketepatan suatu alat ukur dalam mengukur sebuah objek (siswa) sesuai dengan kemampuan yang dimiliki objek tersebut.

Nitko dan Brookhart (2007:67) *reliability is the degree to which students' results remain consistent over replications of an assessment procedure.* Miller (2008:87) *reliability is an estimate of test consistency.* Mardapi (2012:51) reliabilitas atau keandalan merupakan koefisien yang menunjukkan tingkat keajegan atau konsistensi hasil pengukuran suatu tes. Jadi, reliabilitas adalah keajegan alat ukur dalam mengukur kemampuan objek (siswa) dalam kondisi apapun.

Reynolds, Livingston, & Willson (2010:151) menyatakan bahwa "*item discrimination refers to how well an item can accurately discriminate between test takers who differ on the construct being measured*". Menurut Hambleton & Swaminathan, (1985:36) indeks daya beda berada di antara 0 – 2.

Reynolds, Livingston, & Wilson (2010:148) menyatakan bahwa "*item difficulty is defined as the percentage or proportion of test takers who correctly answer the item*". Butir soal yang baik terletak pada interval $-2 \leq \theta \leq 2$ (Hambleton, Swaminathan, & Rogers, 1991:13). Nilai b mendekati -2 maka mengidentifikasi bahwa butir tersebut semakin mudah, dan jika nilai b mendekati +2 maka mengidentifikasi bahwa butir soal semakin sulit.

Parameter c menunjukkan peluang peserta tes dengan kemampuan rendah mampu menjawab butir dengan benar. Pada suatu tes pilihan ganda yang terdiri dari k pilihan alternatif jawaban, maka besarnya parameter c terletak sekitar $\frac{1}{k}$. (Hambleton & Swaminathan 1985:38).

Selanjutnya, karakteristik yang harus dipenuhi oleh perangkat tes adalah kesalahan baku pengukuran harus seminimal mungkin. Menurut Crocker dan Algina (2008:128) *the standard error of measurement is defined as $\sigma_E = \sigma_X \sqrt{1 - \rho_{XX'}}$, and may be considered as the average standard deviation of examinees' individual error distributions for a large number of repeated testings.* Miller (2008:93), *standard error of measurement is a quantitative expression of the magnitude of error in a test score based on the test reliability.*

Studi ini dilakukan untuk mengetahui estimasi kesalahan baku pengukuran pada soal ujian akhir semester akhir matematika kelas XII di SMA Pondok Modern Selamat. Data yang digunakan dalam studi ini

berasal dari lembar jawaban siswa dari soal ujian akhir semester akhir matematika kelas XII di SMA Pondok Modern Selamat yang terdiri dari 120 siswa. Soal ujian akhir semester akhir matematika kelas XII di SMA Pondok Modern Selamat terdiri dari 40 soal yang berbentuk pilihan ganda. Analisis yang dilakukan dalam studi ini adalah analisis karakteristik soal dengan berdasarkan teori tes klasik kemudian analisis estimasi kesalahan baku pengukuran dengan Metode Thorndike, Metode Compound Binomial, Metode Compound Binomial modifikasi Keats, dan Metode ANAVA.

Langkah yang dilakukan untuk mengestimasi kesalahan pengukuran dengan Metode Thorndike dilakukan dengan formula Teori Tes Klasik yaitu dengan membagi perangkat soal menjadi dua bagian sama panjang. Selanjutnya pada masing-masing belahan dicari nilai rerata dan nilai varians. Kemudian menghitung nilai varians kesalahan pada masing-masing belahan dengan menggunakan rumus berikut ini:

$$\sigma_{E1}^2 = \sigma_1^2 \sqrt{1 - r_{xx}}$$

$$\sigma_{E2}^2 = \sigma_2^2 \sqrt{1 - r_{xx}}$$

Setelah nilai varians kesalahan pada masing-masing belahan diperoleh selanjutnya mengestimasi kesalahan pengukuran dengan Metode Thorndike dengan menggunakan rumus berikut ini:

$$\sigma_E = \sqrt{\sigma_{E1}^2 + \sigma_{E2}^2}$$

Metode *Compound Binomial* dilakukan dengan membagi soal UAS menjadi beberapa belahan. Pada metode ini pembagian belahan soal didasarkan pada tingkat kesukaran butir soal. Perhitungan tingkat kesukaran soal dilakukan dengan menggunakan program *Microcat Iteman*. Selanjutnya menghitung nilai varians kesalahan pada masing-masing belahan. Langkah selanjutnya adalah mengestimasi kesalahan baku pengukuran dengan menggunakan Metode *Compound Binomial* dengan rumus berikut ini:

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

Keterangan:

$S_{E(i)}$: 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

Feldt, Steffan, Gupta (1985:353) *Lord conceived of a universe of acceptable test items from which a randomly selected set of size k constitutes a test form*. Gagasan mendasar dalam menentukan besarnya kesalahan baku pengukuran berdasarkan pendekatan binomial yang dikembangkan oleh Lord adalah besarnya standar deviasi nilai yang diperoleh oleh seorang peserta tes pada beberapa tes paralel. Keats memodifikasi formula Lord dengan mengalikan sisi kiri dengan suatu konstanta sehingga menjadi formula sebagai berikut.

$$S_E = \left[\left(\frac{X(k - X)}{k - 1} \right) \left(\frac{1 - r_{xx}'}{1 - r_{21}} \right) \right]^{1/2}$$

Keterangan

S_E : Kesalahan pengukuran bagi peserta tes

X : skor yang diperoleh peserta tes ke i pada rumpun atau kelompok dari butir tes pada kategori h pada sebuah spesifikasi perangkat tes.

C : jumlah kategori butir tes

k : jumlah butir soal

r_{xx}' : reliabilitas perangkat tes
 r_{21} : korelasi perangkat tes kedua dan ketiga

Nilai konstanta masing-masing belahan dapat menggunakan formula KR-21. Formula KR-21 sebagai berikut

$$KR - 21 = \left(\frac{k}{k-1} \right) \left(1 - \frac{k\bar{p}(1-\bar{p})}{s_x^2} \right)$$

Keterangan

K : banyaknya butir soal
 $\bar{p}$: rata-rata tingkat kesukaran
 s_x^2 : varians skor tes

Menurut Feldt, Steffan, Gupta (1985:354) Asumsi dasar yang digunakan oleh Hoyt dalam mengestimasi kesalahan baku pengukuran menggunakan Metode Analisis Varians adalah bahwa "*the examinees by items score matrix for a test may be analyzed to obtain mean squares for examinees (MS_S) items (MS_I) dan interaction (MS_{Sxl}) and the reliability of the test may be estimated from these mean squares*". Hoyt mengemukakan rumus untuk menentukan kesalahan baku pengukuran sebagai berikut:

$$S_E = (k(MS_{Sxl}))^{\frac{1}{2}}$$

Keterangan rumus:

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

Berdasarkan hasil penelitian dan menggunakan langkah-langkah di atas diperoleh hasil estimasi dari masing-masing metode. Hasil estimasi kesalahan pengukuran UAS mata pelajaran matematika SMA Pondok Modern Selamat dengan menggunakan metode thorndike adalah sebesar 2,370, dengan menggunakan metode Compound Binomial adalah sebesar 0,752, dengan menggunakan metode Compound Binomial dengan modifikasi Keats adalah sebesar 1,207, dan dengan menggunakan metode Anava adalah sebesar 2,21.

SIMPULAN

Berdasarkan hasil penelitian dan pembahasan, diperoleh beberapa kesimpulan sebagai berikut:

- 1) Karakteristik soal yang diujikan pada UAS mata pelajaran matematika SMA Pondok Modern Selamat, untuk reliabilitas sebesar 0.786, untuk rata-rata tingkat kesukaran soal sebesar 0.451, sedangkan untuk rata-rata daya beda sebesar 0.423
- 2) Hasil estimasi kesalahan pengukuran UAS mata pelajaran matematika SMA Pondok Modern Selamat dengan menggunakan metode thorndike adalah sebesar 2,370, dengan menggunakan metode Compound Binomial adalah sebesar 0,752, dengan menggunakan metode Compound Binomial dengan modifikasi Keats adalah sebesar 1,207, dan dengan menggunakan metode Anava adalah sebesar 2,21.
- 3) Metode yang paling akurat dalam mengestimasi besarnya kesalahan baku pengukuran soal-soal UAS mata pelajaran Matematika di SMA Pondok Modern Selamat Tahun Ajaran 2018/2019 adalah metode Compound Binomial, karena memberikan hasil estimasi kesalahan terkecil.

REFERENSI

- Azwar, Saifuddin. (2013). *Reliabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar.
- Crocker, Linda., & Algina, James. (2008). *Introduction to Classical and Modern Test Theory*. Mason Ohio: Holt, Rinehart And Winston, Inc.
- Feldt, L. S., Steffen, M., & Gupta, N. C. (1985). *A Comparison of Five Methods for Estimating the Standard Error of Measurement at Specific Score Levels*. Sage Journals, Volume 9, Issue 4 (351-361).
- Hambleton, Ronald. K., & Swaminathan, H. (1985). *Item Response Theory*. Boston: Kluwer, Inc.
- Mardapi, Djemari. (2008). *Teknik Penyusunan Instrumen Tes dan Non Tes*. Yogyakarta: Mitra Cendikia Pres.
- Mardapi, Djemari. (2012). *Pengukuran Penilaian dan Evaluasi Pendidikan*. Yogyakarta: Nuha Medika.
- Miller, M. David., Linn, Robert L., & Gronlund, Norman E. (2009). *Measurement and Assessment in Teaching*. (10th ed). Upper Saddle River, New Jersey: Pearson Education, Inc.
- Nitko, A.J. & Brookhart, S.M. (2007). *Educational Assessment of Students (5th Edition)*. Upper Saddle River, NJ: Person Merrill Prentice Hall.
- Reynolds, C. R., Livingston, R., & Willson, V. (2010). *Measurement and Assessment in Education*. Upper Saddle River, NJ; London: Person Education International.
- Wright, B.D. (2008). *Best Test Design*. Chicago: Mesa Press.

