

## **Estimasi Kesalahan Pengukuran Pendekatan Teori Tes Klasik dan Teori Respon Butir Soal *Tryout* SBMPTN TPA**

Indriani H. Ismail  
Universitas Negeri Makassar, Makassar  
[indrianih.ismail\\_xiia1@yahoo.com](mailto:indrianih.ismail_xiia1@yahoo.com)

**Abstract:** *This research aims to estimate standard error of measurement with CTT and IRT approach and compare the most sensitive theory in estimating standard error of measurement about Tryout SBMPTN TPA Year 2018 in Makassar City. This research is descriptive - explorative research with quantitative approach. The subjects of this study were participants of SBMPTN tryout in 2018 in Makassar city as many as 455 examinee. Data source in this research is answer sheet of SBMPTN TPA trial tryout participants in Makassar City. The instrument used in this study, which is a matter of multiple choice of Academic Potential Test (TPA) consisting of 50 items. The result of standard error estimation based on CTT is  $SE.Meas = 2.95$ ,  $SE.Est = 2.322$  and  $SE.Pred = 3.743$ . While measurement error based on IRT is 0,349. IRT is more sensitive in estimating measurement error about Tryout SBMPTN TPA 2018 in Makassar city.*

**Keywords:** *CTT, IRT, and Standar error of measurement*

**Abstrak:** Penelitian ini bertujuan untuk mengestimasi kesalahan pengukuran dengan pendekatan CTT dan IRT dan membandingkan teori yang paling sensitif dalam mengestimasi kesalahan pengukuran soal *Tryout* SBMPTN TPA Tahun 2018 di Kota Makassar. Penelitian ini merupakan penelitian deskriptif - eksploratif dengan pendekatan kuantitatif. Subyek penelitian ini adalah peserta *tryout* SBMPTN tahun 2018 di kota Makassar sebanyak 455 peserta tes. Sumber data dalam penelitian ini adalah lembar jawaban peserta tes *tryout* SBMPTN TPA di Kota Makassar. Instrumen yang digunakan dalam penelitian ini, yaitu soal pilihan ganda Tes Potensial Akademik (TPA) yang terdiri dari 50 butir soal. Hasil estimasi kesalahan pengukuran berdasarkan CTT yaitu  $SE.Meas = 2.95$ ,  $SE.Est = 2.322$  dan  $SE.Pred = 3.743$ . Sedangkan kesalahan pengukuran berdasarkan IRT yaitu 0,349. IRT lebih sensitif dalam mengestimasi kesalahan pengukuran soal *Tryout* SBMPTN TPA tahun 2018 di kota Makassar.

**Kata Kunci:** CTT, IRT, dan kesalahan pengukuran

## 1. Pendahuluan

Evaluasi merupakan kegiatan penting dalam aktivitas pendidikan terutama dalam proses pembelajaran. Proses pembelajaran tidak hanya membutuhkan pendidik yang mampu mengajar dengan baik, namun juga membutuhkan pendidik yang mampu melakukan evaluasi dengan baik. Dalam undang-undang guru dan dosen pasal 20 ada tujuh tugas utama guru salah satunya adalah melakukan evaluasi terhadap peserta didik.

Ada empat komponen yang penting dalam evaluasi, yaitu evaluasi, penilaian, pengukuran, dan tes atau nontes. Keempat komponen tersebut merupakan satu kesatuan yang saling terkait dan tidak bisa dipisahkan. Evaluasi didahului dengan pengukuran kemudian penilaian. Ketika memulai menyusun dan memberikan tes/nontes kepada responden, maka disaat itulah kita memulai tahapan pengukuran, penilaian dan evaluasi. Ketika kita telah melaksanakan penilaian berarti kita telah melaksanakan pengukuran.

Dengan melakukan pengukuran, pendidik akan mengetahui tingkat kemampuan peserta didik dalam mata pelajaran tertentu. Pengukuran dalam pendidikan tidak dapat dipisahkan dari pengukuran kemampuan peserta didik. Pengukuran dalam pendidikan meliputi pengukuran kemampuan peserta tes dan pengukuran karakteristik alat ukur yang digunakan.

Setiap pengukuran selalu mengandung kesalahan yang disebut dengan *standar error of measurement* (SEM). Sumber kesalahan pengukuran diantaranya adalah pada penentuan materi ujian, pihak yang diukur, pihak yang mengukur, dan lingkungan. Keadaan kesehatan fisik dan emosi

peserta didik yang bervariasi dari waktu ke waktu. Untuk mengatasi kesalahan pada objek yang diukur, sebaiknya sering melakukan pengukuran, sedangkan untuk mengatasi kesalahan pada pihak yang mengukur, ia harus dilatih agar mampu menyusun alat ukur dengan baik dan mampu menyelenggarakan pengukuran dengan kondisi yang standar.

Wright (2008:130) menyatakan bahwa  $\text{True score} = \text{observed score} \pm \text{measurement error}$ . (skor sebenarnya = skor hasil pengukuran  $\pm$  kesalahan pengukuran). 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.

Ada dua teori pengukuran yang sampai saat ini masih dikembangkan, yaitu teori tes klasik dan teori tes modern. Teori tes klasik disebut juga *classical test theory* (CTT), sedang teori tes modern disebut juga *item response theory* (IRT).

Prosedur analisis dan seleksi item soal berdasarkan teori tes klasik memperhatikan tiga parameter, yaitu (1) tingkat kesukaran item, (2) daya diskriminasi item, dan (3) efektivitas distraktor (Azwar, 2015). Terdapat tiga asumsi dasar dalam Teori Respon Butir yang harus dipenuhi. Asumsi-asumsi tersebut adalah unidimensi, independensi lokal, dan ketepatan kurva karakteristik butir (Hambleton, dkk, 1991).

Model teori respon butir berdasarkan jumlah parameter butir, yaitu 1-P, 2-P, dan 3-P. Peluang

menjawab benar suatu butir soal sebagai berikut :

$$3\text{-P } P_i(\theta) = c_i + (1 - c_i) \left( \frac{e^{1.7a_i(\theta-b)}}{1 + e^{1.7a_i(\theta-b)}} \right) \dots (1)$$

$$2\text{-P } P_i(\theta) = \left( \frac{e^{1.7a_i(\theta-b)}}{1 + e^{1.7a_i(\theta-b)}} \right) \dots (2)$$

$$1\text{-P } P_i(\theta) = \left( \frac{e^{1.7(\theta-b)}}{1 + e^{1.7(\theta-b)}} \right) \dots (3)$$

Dimana a adalah daya beda, b adalah tingkat kesukaran, c adalah pseudo gussing, e adalah eksponen dan  $\theta$  adalah kemampuan (Mardapi, 2012).

Secara konseptual, kesalahan baku pengukuran terkait dengan uji reliabilitas karena itu memberikan indikasi dispersi kesalahan pengukuran ketika Anda mencoba untuk memperkirakan nilai sebenarnya siswa dari nilai tes yang diamati. Untuk memahami kalimat sebelumnya, Anda harus terlebih dahulu memahami tiga istilah: kesalahan pengambilan sampel, skor sebenarnya, dan skor tes. Kesalahan baku pengukuran dapat dihitung dari standard deviasi skor tes dan reliabilitas dengan rumus sebagai berikut :

$$SEM = s\sqrt{1 - r_{xx}} \dots (4)$$

Dimana, SEM adalah *standard error of measurement*, s adalah standar deviasi dan  $r_{xx}$  adalah reliabilitas (Brown, 1999).

Pada teori respon butir dikenal istilah fungsi informasi. Fungsi informasi merupakan kecocokan di antara parameter kemampuan dengan parameter butir maka kekeliruan baku merupakan ketidakcocokannya. Pada penentuan fungsi informasi, telah digunakan variansi yakni kuadrat dari simpangan baku. Kekeliruan baku adalah kebalikan dari fungsi informasi butir pada besaran simpangan bakunya. Karena itu kekeliruan baku merupakan

kebalikan dari akar fungsi informasi ujian dengan persamaan :

$$KB(\theta) = \frac{1}{\sqrt{I(\theta)}} \dots (5)$$

Makin tinggi fungsi informasi makin kecil kekeliruan baku. Sebaliknya makin rendah fungsi informasi makin besar kekeliruan baku (Naga, 2012).

Ada dua macam kesalahan pengukuran, yaitu kesalahan acak dan kesalahan sistematis (Mardapi, 2008:3). Kesalahan acak disebabkan oleh kondisi fisik dan mental yang diukur dan mengukur bervariasi. Kesalahan sistematis terjadi karena alat ukur yang digunakan.

Dewasa ini, kebutuhan masyarakat Indonesia akan pendidikan sangat tinggi. Masyarakat Indonesia sangat antusias untuk melanjutkan jenjang pendidikan dari Sekolah Menengah Atas atau kejuruan ke tahap pendidikan yang lebih tinggi yang diselenggarakan oleh Perguruan Tinggi Negeri atau swasta. Berdasarkan Undang-Undang Nomor 12 Tahun 2012 tentang Pendidikan Tinggi. Peraturan Pemerintah Nomor 4 Tahun 2014 tentang penyelenggaraan Pendidikan Tinggi dan Pengelolaan Perguruan Tinggi, dan Peraturan Menteri Riset, Teknologi, dan Pendidikan Tinggi Nasional Nomor 2 Tahun 2015 tentang Penerimaan Mahasiswa Baru Program Sarjana pada Perguruan Tinggi Negeri, pola penerimaan mahasiswa baru program sarjana pada perguruan tinggi dilakukan melalui: Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN); Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN); dan Penerimaan Jalur Mandiri.

Tes potensi merupakan salah-satu bentuk pengukuran terhadap kemampuan abilitas kognitif potensial umum (pengukuran performansi

maksimal) yang dirancang khusus guna memprediksi peluang keberhasilan belajar di perguruan tinggi, karena itulah tes seperti ini biasanya dinamai Tes Potensi Akademik.

Sebelum mengikuti seleksi SBMPTN banyak siswa mengikuti bimbingan belajar disuatu lembaga bimbingan belajar. Metode pembelajaran pada bimbingan belajar yaitu calon mahasiswa diperkenalkan dan diajarkan cara menyelesaikan soal-soal SBMPTN dengan teknik yang cepat atau cara pintas dalam menjawab soal SBMPTN. Selain itu banyak lembaga bimbingan belajar juga melaksanakan *tryout* SBMPTN sebagai ajang latihan untuk calon mahasiswa sebelum menghadapi SBMPTN. Permasalahannya adalah perangkat soal tersebut belum pernah dilakukan analisis butir soal untuk menguji validitas dan reliabilitas suatu perangkat tes. Sehubungan dengan penjelasan di atas maka perlu untuk dilakukan suatu perhitungan estimasi kesalahan dalam suatu pengukuran yang disebut dengan estimasi kesalahan pengukuran atau *standard error of measurement* (SEM). Ada berbagai metode yang dapat digunakan untuk menghitung estimasi kesalahan pengukuran, yaitu ada metode berdasarkan teori klasik dan ada metode berdasarkan teori respon butir.

Hasil penelitian Wahyuni dan Mardapi (2015) telah melakukan penelitian mengenai komparasi beberapa metode estimasi kesalahan pengukuran soal ujian akhir semester matematika SMA di kabupaten Lampung Tengah yaitu : hasil analisis data menunjukkan bahwa besarnya estimasi kesalahan pengukuran perangkat soal ujian akhir semester matematika SMA di kabupaten

Lampung Tengah dengan menggunakan metode Thordike, metode compound binomial, metode polynomial, dan metode Teori Respon Butir. Metode yang paling akurat dalam mengestimasi kesalahan pengukuran adalah Metode *Compound Binomial*.

Kharismawati (2013) mengestimasi kesalahan bakupengukuran soal ujian akhir semester bahasa inggris sekolah menengah pertama di Daerah Istimewa Yogyakarta. Metode yang digunakan adalah Metode Teori Tes Klasik, pendekatan Binomial Lord dengan Modifikasi Keat, Metode Compoun Binomial, Metode ANOVA, formula umum Teori Tes Modern dan Item Respons Curve. Hasil estimasi tersebut menunjukkan bahwa, estimasi kesalahan baku pengukuran dengan pendekatan teori tes klasik dengan menggunakan Metode Compound Binomial memberi estimasi kesalahan baku pengukuran paling kecil.

Andrian (2014) meneliti tentang Estimasi kesalahan bakupengukuran soal-soal uas matematika kelas XII SMA N di kota selat panjang kabupaten kepulauan meranti Provinsi riau. Estimasi Kesalahan Pengukuran soa-soal ujian akhir semester Matemati kakelas XII SMAN/ sederajat di Kota Selatpanjang Kabupaten Kepulauan Meranti Provinsi Riau dengan tiga metode estimasi berdasarkan teori tes klasik. Metode yang digunakan antara lain Metode Thorndike, Metode Compound Binomial dan Metode Analisis Varians dapat disimpulkan metode yang paling tepat untuk mengestimasi kesalahan pengukuran adalah Metode Compound Binomial.

Hasil penelitian Prijowuntato (2015) mengenai perbandingan

estimasi kesalahan pengukuran *standard setting* dalam penilaian kompetensi akuntansi SMK yaitu metode angoff merupakan metode yang tepat untuk mengestimasi kesalahan pengukuran pada *standard setting*.

## 2. Metode Penelitian

Penelitian ini termasuk penelitian deskriptif-eksploratif dengan pendekatan kuantitatif. Subyek penelitian ini adalah peserta tryout SBMPTN tahun 2018 di kota Makassar sebanyak 455 peserta tes. Sumber data dalam penelitian ini adalah lembar jawaban peserta tryout SBMPTN TPA. Lembar jawaban peserta didik kemudian dianalisis untuk mengestimasi kesalahan baku pengukuran dari perangkat tes tryout SBMPTN Tahun 2018 Mata Pelajaran TPA di Kota Makassar. Instrumen yang digunakan dalam penelitian ini, yaitu soal pilihan ganda Tes Potensial Akademik (TPA) yang terdiri dari 50 butir soal.

## 3. Hasil Penelitian dan Pembahasan

Analisis kuantitatif soal tryout SBMPTN TPA berdasarkan Teori Tes Klasik dilakukan dengan menggunakan program R . Hasil yang diperoleh dari Program R adalah adalah tingkat kesukaran (b), daya beda (a), efektivitas distraktor (pengecoh). Selain informasi karakteristik butir soal, informasi lain yang diperoleh adalah hasil perhitungan deskriptif berupa jumlah keseluruhan butir soal, jumlah peserta tes, rerata (*mean*), reliabilitas, dan standar deviasi.

Setelah analisis butir soal berdasarkan Teori Tes Klasik juga dilakukan analisis berdasarkan Teori Respon Butir. Ada beberapa analisis yang dilakukan dalam metode Teori Respon Butir yaitu uji parameter logistik (1PL, 2PL dan 3PL), uji kecocokan model, kurva karakteristik butir, fungsi informasi dan kesalahan baku pengukuran sebagai *standar error of measurement (SEM)*.

**Tabel 1.** Hasil Analisis data berdasarkan Teori Tes Klasik

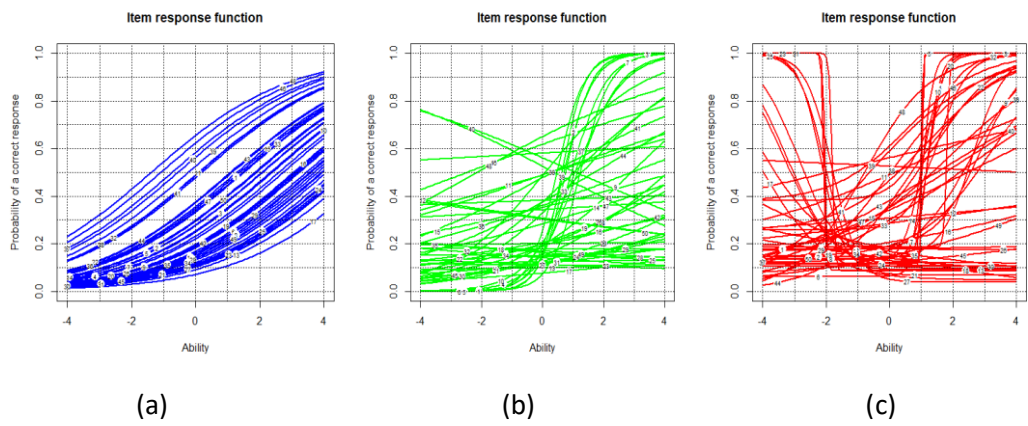
Analisis	Nilai
Jumlah Responden	455
Jumlah Butir Soal	50
Mean	14,057
Standar deviasi	4,794
Reliabilitas	0,626

**Tabel 2.** Standard error dan Standard error of Measurement Berdasarkan Teori Tes Klasik

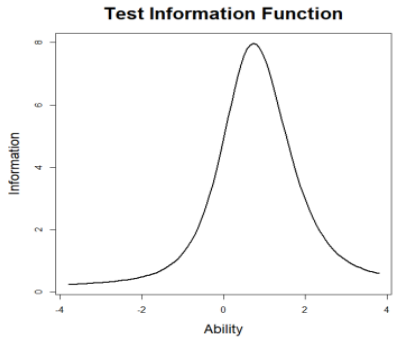
Jenis SE	Nilai
SE Meas.	2,935
SE Est.	2,322
SE Pred.	3,743

**Tabel 3.** Rangkuman hasil uji kecocokan model dengan Teori Respon Butir

Model Parameter Logistik	Jumlah Item Fit
1PL	34
2PL	39
3PL	18



Gambar 1. Kurva Karakteristik (a) 1PL, (b) 2PL, dan (c) 3PL



Gambar 2. Kurva Karakteristik (a) 1PL, (b) 2PL, dan (c) 3PL

Tabel 4. Kesalahan Baku Pengukuran dengan Teori Respon Butir

Model Logistik	Teta ( $\theta$ )	SEM		
		Min	Maks	SEM TEST
2 PL	$-4 \leq \theta \leq 4$	0,471	12,867	0,349

Nilai kesalahan pengukuran dapat digunakan untuk mengestimasi interval skor sebenarnya (*true score*) pada masing-masing peserta didik. Hal ini didasarkan pada Teori Tes Klasik bahwa  $X = T + E$  dengan X adalah skor tampak atau skor yang diperoleh dari hasil tes, T adalah *true score* atau skor sebenarnya, dan E kesalahan pengukuran berdasarkan Teori Tes Klasik atau *Classical Test Theory* (CTT) dan Teori Respon Butir atau *Item Response Teori* (IRT).

Akan tetapi, tidak semua nilai *standard error* di atas digunakan untuk menentukan nilai *true score* (T). *Standar error of measurement* (SE.Meas) berguna dalam konstruksi taraf kepercayaan tentang *true score*. *Standar error of estimate* (SE.Est) berguna dalam konstruksi taraf kepercayaan tentang skor yang diamati. *Standar error of prediction* (SE.Pre) berguna dalam menghitung *true score*.

Hal ini dijelaskan dalam Dudek (1979) yang menyatakan bahwa :

“Individuals often misinterpret the SEM. In fact, most textbooks misinterpret these measures. The SE.Meas ( $s*\sqrt{1-rxx}$ ) is useful in the construction of CI about observed scores, but should not be interpreted as indicating the TRUE SCORE is necessarily included in the CI. The SE.Est ( $s*\sqrt{rxx*(1-rxx)}$ ) is useful in the construction of CI about the TRUE SCORE. The estimate of a CI for a TRUE SCORE also requires the calculation of a TRUE SCORE (due to regression to the mean) from observed scores. The SE.Pred ( $sy*\sqrt{1-rxx^2}$ ) is useful in predicting the score on a parallel measure (Y) given a score on test X. SE.Pred is usually used to estimate the score of a re-test of an individual.”

Setelah menganalisis kurva karakteristik butir, analisis selanjutnya yaitu uji kecocokan model perangkat tes terhadap suatu model parameter logistik. Uji kecocokan butir bertujuan untuk mengetahui perangkat tes cocok dengan model logistik 1PL, 2PL ataukah 3PL. Berdasarkan hasil analisis kuantitatif karakteristik soal dengan Teori Respon Butir diperoleh kesimpulan bahwa perangkat soal *tryout* SBMPTN TPA memiliki kecocokan model dengan model 2 parameter logistik (2PL).

Analisis data selanjutnya adalah analisis data fungsi informasi soal *Tryout* SBMPTN TPA tahun 2018. Karena perangkat tes soal *Tryout* SBMPTN TPA tahun 2018 cocok dengan model 2PL maka fungsi informasi yang dianalisis pada

penelitian ini adalah fungsi informasi 2PL. Menghitung nilai fungsi informasi pada masing-masing butir pada kemampuan ( $\theta$ )  $-4 \leq \theta \leq 4$  dengan nilai perubahan ( $\theta$ ) sebesar 1.

Nilai kesalahan pengukuran yang diperoleh dengan menggunakan Metode Teori Respon Butir pada perangkat soal *tryout* SBMPTN diperoleh pada rentang kemampuan  $-4 \leq \theta \leq 4$ . Estimasi kesalahan pengukuran yaitu 0,349 dengan estimasi kesalahan pengukuran terkecil 0,471 pada  $\theta = (+1) - (+2)$  dan estimasi kesalahan terbesar 12,877 pada  $\theta = (-4) - (-3)$ .

Hasil estimasi kesalahan pengukuran dengan CTT dan IRT menunjukkan bahwa terdapat variasi harga estimasi kesalahan pengukuran perangkat soal *tryout* SBMPTN TPA tahun 2018 di kota Makassar. Estimasi kesalahan pengukuran terkecil diperoleh berdasarkan IRT. Meskipun asumsi yang digunakan dalam Teori Respon Butir berbeda, akan tetapi dalam penelitian ini estimasi kesalahan pengukuran dengan Teori Respon Butir dilakukan dalam skor mentah yang sama seperti teori tes klasik, sehingga hasilnya dapat dibandingkan secara empirik. Hasil analisis kesalahan pengukuran berdasarkan teori tes klasik sangat tergantung pada karakteristik peserta yang diukur. Apabila perangkat soal tersebut diujikan pada sekelompok peserta yang lain, maka hasil analisisnya akan berubah sesuai dengan karakteristik peserta tesnya. Meskipun demikian, dengan cara ini dapat memberikan rambu-rambu terhadap pembuatan soal yang baik. Kelemahan-kelemahan yang terdapat dalam teori tes klasik dapat diatasi dengan teori respons butir. Teori respons butir melepaskan keterikatan antara butir

soal dengan peserta tes. Karakteristik peserta akan tetap sama walau mengerjakan butir soal yang berbeda, demikian pula karakteristik butir soal juga akan tetap sama walau dikerjakan oleh peserta yang berbeda kemampuannya. Secara ringkas dapat dikatakan bahwa semua metode mempunyai dasar teoritis yang sama. Sebagaimana yang dijelaskan dalam Widayati (2009 : 15).

“Meskipun metode-metode estimasi nampak berbeda, tetapi memiliki variasi dalam konsep yang sama yaitu jika peserta tes dapat dikelompokkan berdasar skor murninya, varians kesalahannya akan sama dengan  $\sum P_j - (1 - P_j)$  dimana  $P_j$  merupakan proporsi menjawab benar untuk item  $j$ . Nampak jelas bahwa kesamaan data empiris untuk beberapa metode mencerminkan kesamaan konsep diantara metode-metode tersebut. Pemilihan berbagai metode tergantung pada pertimbangan praktis dan pada kesukaan pengguna terhadap logika yang melandasi masing-masing metode.“

Meskipun secara sekilas teori tes klasik memiliki banyak keterbatasan, namun bukan berarti pendekatan ini tidak memiliki kelebihan. Beberapa kelebihan teori tes klasik dibandingkan teori respon butir diantaranya adalah terkait dengan analisis yang lebih sederhana dan mudah dipahami. Teori ini memiliki nilai praktis yang tinggi dalam menerangkan masalah reliabilitas dan validitas, disamping pemahamannya yang tidak menuntut pengetahuan mendalam mengenai fungsi distribusi statistik dan model-model matematikanya (Azwar, 2011).

Selain itu dalam melakukan analisis, sampel yang diperlukan bisa lebih sedikit jika dibandingkan dengan pendekatan teori respon butir, sehingga sangat cocok dilakukan untuk penilaian di kelas dan perbaikan evaluasi pembelajaran.

#### 4. Kesimpulan

Hasil estimasi kesalahan pengukuran berdasarkan teori tes klasik yaitu  $SE.Meas = 2.95$ ,  $SE.Est = 2.322$  dan  $SE.Pred = 3.743$ . Sedangkan kesalahan pengukuran berdasarkan teori respon butir yaitu 0,349. Teori respon butir lebih sensitif dalam mengestimasi kesalahan pengukuran soal Tryout SBMPTN TPA tahun 2018 di kota Makassar.

#### 5. Daftar Pustaka

- Azwar, S. (2015). Penyusunan Skala Psikologi Edisi 2. Yogyakarta : Pustaka Pelajar.
- Andrian. (2014). Estimasi Kesalahan Baku Pengukuran Soal-Soal UAS Matematika Kelas XII SMAN di Kota Selat Panjang Kabupaten Kepulauan Meranti Provinsi Riau. Tesis magister, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta.
- Brown, J.D. (1999). Standard error vs. Standard error of measurement. JALT Testing & Evaluation SIG Newsletter, 3(1) : 20-25.
- Dudek, F. J. (1979). The continuing misinterpretation of the standard error of



- measurement. *Psychological Bulletin*, 86, 335-337.
- Feldt, L.S., & Steffen, M., & Gupta C.N. (1985). A comparison of five methods for estimating the standard error of measurement at specific score model. *Applied Psychological Measurement*, 9 (4), 351-361.
- Hambleton, R.K., & Swaminathan H. (1985). *Item Response Theory principles and applications*. New York : Sage Publications.
- Kharismawati, L.R.S. (2013). Estimasi Kesalahan Pengukuran Soal Ujian Akhir Semester Bahasa Inggris Sekolah Menengah Pertama di DIY. Tesis. Tidak diterbitkan. Yogyakarta: Universitas Negeri Yogyakarta,
- Mardapi, D. (2008). Teknik penyusunan instrumen tes & non tes. Yogyakarta: Mitra Cendikia
- (2012). *Pengukuran, Penilaian dan Evaluasi Pendidikan*. Yogyakarta : Nuha Medika.
- Naga, D.S. (2013). *Teori Sekor pada Pengukuran Mental*. 2013. Jakarta : PT Nagarani Citrayasa.
- Prijowuntato, dkk. (2015). Perbandingan Estimasi Kesalahan Pengukuran Standard Setting dalam Penilaian Kompetensi Akuntansi SMK. *Jurnal Penelitian dan Evaluasi Pendidikan*, 19(2), 176-188.
- Wahyuni, S & Mardapi, J. (2015). Komparasi Metode Estimasi Kesalahan Pengukuran Soal Ujian Akhir Semester Matematika SMA di Kabupaten Lampung Tengah. *Jurnal Evaluasi Pendidikan*, 3 (2), 191-201.
- Widayati, C.S.W. (2009). Komparasi beberapa Metode Estimasi Kesalahan Pengukuran. *Jurnal Penelitian dan Evaluasi Pendidikan*, 13 (2), 1-16
- Wright, R. J. (2008) *Educational assessment*. Thousand Oaks: Sage Publications