

Estimasi Kesalahan Pengukuran Berdasarkan CTT dan IRT Soal Tryout SBMPTN TPA Tahun 2018 di Kota Makassar

Estimation of Measurement Errors Based on CTT and IRT of the 2018 SBMPTN Tryout Questions for Academic Potential Tests in Makassar City

Indriani H Ismail

STAI DDI Pangkajene Sidenreng Rappang

*Email: indrianih.ismail_xia1@yahoo.com

ABSTRACT

This research aims to estimate standard error of measurement with TCC 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*

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.

Hasil pengukuran yang berupa angka atau skor diharapkan mencerminkan kemampuan peserta tes yang sebenarnya. Namun demikian ada kemungkinan terdapat perbedaan antara skor hasil pengukuran dengan skor sebenarnya. Ada dua macam kemungkinan, pertama skor hasil pengukuran lebih rendah daripada skor yang sebenarnya. Kedua, skor yang diberikan kepada peserta tes lebih tinggi daripada skor sebenarnya. Bila kemungkinan itu terjadi, berarti terjadi kesalahan pengukuran.

Kegiatan pengukuran merupakan kegiatan yang pertama kali dilakukan dalam melakukan evaluasi. Pengukuran dilakukan dengan menggunakan sebuah instrumen yang telah memiliki karakteristik yang baik agar hasil dari pengukuran benar-benar menggambarkan objek yang diukur. Hasil pengukuran yang cermat akan mencerminkan sebuah informasi yang sebenarnya dari objek yang diukur, misalnya dalam pengukuran hasil belajar peserta didik. Pengukuran harus dilakukan dengan cermat akan memberikan informasi yang diperoleh juga tepat sedangkan pengukuran yang kurang cermat akan memberikan hasil dan informasi yang kurang tepat. Oleh karena itu, hasil pengukuran memberikan kemungkinan bahwa ada perbedaan antara skor hasil dari pengukuran dengan skor sebenarnya. Kemungkinan yang muncul adalah hasil pengukuran lebih besar dari nilai sebenarnya atau sebaliknya, hasil pengukuran lebih kecil dari nilai sebenarnya. Kemungkinan tersebut terjadi karena adanya kesalahan pengukuran.

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.

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). Teori klasik menggunakan asumsi bahwa skor yang diperoleh seseorang dari hasil suatu pengukuran dapat diuraikan menjadi skor yang sebenarnya dan skor kesalahan. Selanjutnya diasumsikan bahwa tidak ada hubungan antara skor yang sebenarnya dengan skor kesalahan. Dengan menggunakan dua asumsi dasar ini selanjutnya dikembangkan formula-formula untuk mengetahui indeks kesahihan (*validity*) dan indeks keandalan (*reliability*).

Teori tes klasik ini memiliki beberapa kelemahan. Kelemahan yang menonjol adalah ketergantungan statistik butir pada karakteristik kelompok yang diukur. Sehingga besarnya statistik butir bervariasi dari satu kelompok terhadap kelompok yang lain. Statistik butir sering juga disebut dengan parameter butir, yaitu tingkat kesulitan dan daya pembeda butir tes. Akibatnya sulit membandingkan kemampuan kelompok satu dengan yang lainnya, apalagi antara individu. Kelemahan ini sudah disadari, yaitu sejak dikembangkan alat ukur yang digunakan pada bidang ilmu-ilmu alam atau teknologi. Alat ukur yang digunakan pada bidang ini tidak bergantung pada objek yang diukur, karena karakteristiknya tidak berubah-ubah selama objek yang diukur sama. Hal ini mudah dipahami karena yang diukur adalah objek yang mati. Lain halnya dengan objek pada bidang pendidikan, yaitu manusia. Keadaan manusia, seperti kondisi susah, senang, selalu berubah-ubah dari waktu ke waktu sehingga hasil pengukuran yang diperoleh belum tentu menunjukkan kemampuan individu yang sebenarnya. Oleh karena itu, dikembangkan teori pengukuran yang dapat mengatasi kelemahan teori klasik.

Teori pengukuran yang berkembang saat ini disebut dengan teori modern atau teori respon butir. Asumsi utama teori ini adalah peluang seseorang menjawab benar suatu butir tidak ditentukan oleh peluang menjawab benar butir yang lain, dikenal dengan asumsi independen. Teori modern berusaha untuk mengembangkan suatu analisis yang menghasilkan estimasi kemampuan seseorang tanpa dipengaruhi alat ukur yang digunakan. Estimasi kemampuan peserta didik tidak berdasarkan jumlah butir soal yang dijawab benar, tetapi berdasarkan pada pola respons dan jumlah butir yang dijawab benar. Demikian juga statistik butir juga diusahakan agar tidak tergantung pada karakteristik individu yang diukur. Berdasarkan sifat-sifat ini, penerapan teori tes modern yang dikembangkan oleh banyak pakar pengukur.

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 Widayanti (2009) telah melakukan penelitian mengenai komporasi beberapa metode estimasi kesalahan pengukuran yaitu : hasil analisis data menunjukkan bahwa besarnya estimasi kesalahan pengukuran perangkat soal tes ujicoba ujian nasional mata pelajaran biologi SMA tahun pelajaran 2007/2008 dengan menggunakan metode Thordike, Polynomial, Binomial dengan modifikasi Keats, Compound Binomial, Analisis Varians, dan teori respon butir. Hasil perhitungan *root mean square error* (RMSE) menunjukkan bahwa metode teori respon butir merupakan metode yang paling tepat dalam menghitung estimasi kesalahan pengukuran.

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*.

METODE

Penelitian ini termasuk penelitian deskriptif-eksploratif dengan pendekatan kuantitatif. Subyek penelitian ini adalah peserta tryout SBMPTN tahun 2018 di kota Makassar yang diadakan oleh bimbingan belajar Focus A sebanyak 455 peserta tes. Sumber data dalam penelitian ini adalah lembar jawaban peserta tryout SBMPTN Mata Pelajaran TPA di Kota Makassar. 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.

HASIL DAN PEMBAHASAN

Analisis kuantitatif soal *tryout* SBMPTN TPA berdasarkan Teori Tes Klasik dilakukan dengan menggunakan program R. Hasil yang diperoleh dari Program R adalah tingkat kesukaran (b), daya beda (a), efektivitas distraktor (pengecoh), validitas dan reliabilitas.

Analisis kuantitatif karakteristik soal *tryout* SBMPTN TPA didasarkan pada parameter tingkat kesukaran, daya beda, dan efektifitas distraktor. Ketiga parameter ini terdiri dari beberapa kriteria berdasarkan hasil analisis dengan program R. Rangkuman hasil analisis kuantitatif parameter daya beda, indeks kesukaran dan validitas dengan program R dapat dilihat pada Tabel 1.

Tabel 1. Distribusi Kategori Parameter tingkat kesukaran, daya beda, dan validitas Hasil Analisis *Program R*

Parameter	Kriteria	Frekuensi
Tingkat Kesukaran	Mudah	2 butir
	Sedang	18 butir
	Sukar	30 butir
Daya Beda	Jelek	10 butir
	Cukup	25 butir
	Baik	15 butir
	Valid	38 butir
Validitas	tidak valid	12 butir

Selain informasi karakteristik butir soal, informasi lain yang diperoleh menggunakan Program R adalah hasil perhitungan deskriptif berupa jumlah keseluruhan butir soal, jumlah peserta tes, rerata (*mean*), reliabilitas, dan standar deviasi. Hasil analisis statistik deskriptif perangkat soal *tryout* SBMPTN TPA dapat dilihat pada Tabel 2.

Tabel 2. Reliabilitas tes hasil Program R

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

Selain analisis kuantitatif soal berdasarkan Teori Tes Klasik juga dilakukan analisis berdasarkan Teori Respon Butir. Ada beberapa analisis yang dilakukan dalam metode Teori Respon Butir yaitu uji prasyarat (unidimensi, invarian, dan independensi lokal), pola responden, uji parameter logistik (1PL, 2PL dan 3PL), uji kecocokan model, kurva karakteristik butir, uji kecocokan butir (*item fit*), fungsi informasi dan kesalahan baku pengukuran sebagai *standar error of measurement* (SEM). Analisis kuantitatif berdasarkan Teori Respon Butir dilakukan untuk melihat model yang cocok dengan perangkat soal *tryout* SBMPTN TPA di kota Makassar. Analisis kuantitatif karakteristik soal berdasarkan Teori Respon Butir dilakukan dengan menggunakan Program R.

Tabel 3. Distribusi Kategori Parameter Butir Model 1PL

Tingkat Kesukaran (b)		
Keterangan	Frekuensi	Persentase (%)
baik	16	32
tidak baik	34	68
Total	50	100

Tabel 4. Distribusi Kategori Parameter Butir Model 2PL

Daya Beda (a)		
Keterangan	Frekuensi	Persentase (%)
baik	12	24
tidak baik	38	76
Total	50	100

Tingkat Kesukaran (b)		
Keterangan	Frekuensi	Persentase (%)
baik	38	76
tidak baik	12	24
Total	50	100

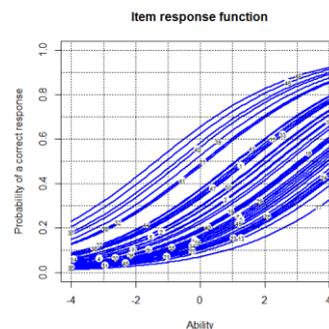
Tabel 5. Distribusi Kategori Parameter Butir Model 3PL

Daya Beda (a)		
Keterangan	Frekuensi	Persentase (%)
baik	11	22
tidak baik	39	78

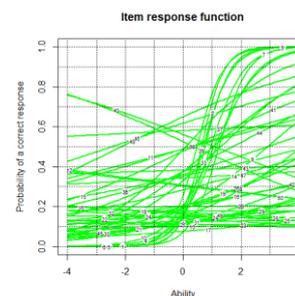
Tingkat Kesukaran (b)		
Keterangan	Frekuensi	Persentase (%)
baik	18	36
tidak baik	32	64
Total	50	100

Guessing ©		
Keterangan	Frekuensi	Persentase (%)
baik	44	88
tidak baik	6	12
Total	50	100

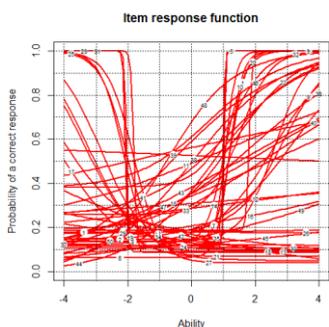
Karakteristik butir soal yang terdiri dari tingkat kesukaran (b), daya pembeda (a), dan tebakan atau *guessing* (c) juga dapat dilihat dalam suatu kurva karakteristik butir atau *item characteristic curve* (ICC).



Gambar 1. Kurva karakteristik butir 1PL soal *tryout* SBMPTN TPA



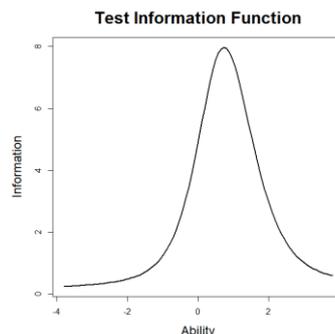
Gambar 2. Kurva karakteristik butir 2PL soal *tryout* SBMPTN TPA



Gambar 3. Kurva karakteristik butir 3PL soal tryout SBMPTN TPA

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 tahun 2018 di kota Makassar 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 (θ) $-4 \leq \theta \leq 4$ dengan nilai perubahan (θ) sebesar 1.



Gambar 4. kurva fungsi informasi tes tryout SBMPTN TPA Tahun 2018

Hasil analisis standar error (SE) berdasarkan Teori Tes Klasik dapat dilihat pada Tabel 6 di bawah ini.

Tabel 6. SE Berdasarkan Teori Klasik

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

Setelah mengetahui model parameter logistik yang cocok dengan perangkat tes dan nilai fungsi informasinya, maka dapat dilakukan estimasi kesalahan pengukuran (*Standar Error*). Pada Tabel 7 dapat dilihat hasil perhitungan estimasi kesalahan pengukuran soal tryout SBMPTN TPA Tahun 2018 di kota Makassar.

Tabel 7. SE Berdasarkan Metode IRT

Model Logistik	Teta (θ)	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).

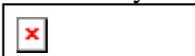
Akan tetapi, tidak semua nilai *standard error* di atas digunakan untuk menentukan nilai *true score* (T). *Standar error of measurement* (SE.Meas) berguna dalam konstruksi taraf kepercayaan tentang *true score*. *Standar error of estimate* (SE.Est) berguna dalam konstruksi taraf kepercayaan tentang 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 ($s \cdot \sqrt{1-r_{xx}}$) 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 ($s \cdot \sqrt{r_{xx} \cdot (1-r_{xx})}$) 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 ($s_y \cdot \sqrt{1-r_{xx}^2}$) 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.”

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 Widiyati (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  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.“

SIMPULAN

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.

DAFTAR RUJUKAN

- 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.
- Allen, M.J. & Yen, W.M. 1979. *Introduction to measurement theory*. Monterey, CA: Books/Cole Publishing Company.
- Dodi Arianto (2009). Estimasi kesalahan bakupengukuran soal-soal Matematika IX ulangan akhir semester (UAS) 1 SMP di kota Yogyakarta.
- Azwar, S. 2000. *Reliabilitas dan validitas (edisi 3)*. Yogyakarta: Pustaka Pelajar.
- , 2005. Tes Prestasi. *Fungsi dan Pengembangan Pengukuran Prestasi Belajar*. Yogyakarta : Pustaka Pelajar Offset.
- , 2007. *Reliabilitas dan validitas (edisi ke-4)*. Yogyakarta : Pusaka Pelajar.
- Crocker, L., & Algina, J. 2008. *Introduction to Classical and Modern Theory*. New York: CBS - College Publishing
- 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.
- Hadi, S. 2014. *Lembar Kerja Teknik Analisis Data Pascasarjana, Mengolah hasil analisis Bilog dengan SPSS*. Materi Kuliah Praktek Pengukuran dan Pengujian, Pascasarjana UNY.
- 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. 1999. Estimasi Kesalahan Pengukuran dalam Bidang Pendidikan dan implikasinya pada Ujian Nasional. Pidato Pengukuhan Guru Besar Madya Ilmu Evaluasi Pendidikan Teknologi dan Keguruan pada Fakultas Teknik Universitas Negeri Yogyakarta.
- , 2012. *Pengukuran, Penilaian dan Evaluasi Pendidikan*. Yogyakarta : Nuha Medika.
- Mansyur, dkk. 2015. *Asesmen Pembelajaran di Sekolah*. Yogyakarta: Pustaka Pelajar.
- Mehrens, W.A & Lehmann, J.J (1973). *Measurement and evaluation in education and psychology*. New York : Hold, Rinehart and Wiston, Inc
- 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.
- Ramadhan, S. 2015. Estimasi Kesalahan Baku Pengukuran Soal – Soal UAS Fisika Kelas XII SMA di Kabupaten Bima Provinsi NTT. Yogyakarta: Pascasarjana UNY
- Ruslan, R. 2017. *Estimasi Kesalahan Baku Pengukuran Soal Tryout USBN Kimia SMA Kota Makassar*. Makassar : Pascasarjana UNM.
- Sekaran,U. 2006. *Metodelogi Penelitian untuk bisnis*. Jakarta : Salemba Empat.
- Sharkness,J. & De Angelo.L. 2010. Measuring theory in Language testing: Past tradiditions and current trends. *I-Manager's Journal on Educational Pscology*,3 (2), 1-12.
- Sudaryono, dkk. 2013. *Pengembangan Instrumen Penelitian Pendidikan*. Yogyakarta: Graha Ilmu.
- Sudijono,A. 2005. *Pengantar Evaluasi Pendidikan*. Jakarta : Raja Grafindo Persada.
- Sukardi. 2008. *Evaluasi Pendidikan*. Jakarta: Bumi Aksara
- Susetyo. 2015. *Prosedur Penyusunan & Analisis Tes*. Bandung : PT Refika Aditama.
- Surapranata, S. 2005. *Analisis, Validitas, Reliabilitas dan Interpretasi Hasil Tes*. Implementasi Kurikulum 2004. Bandung: Remaja Rosdakarya Offset.
- Suryabrata, S.1982. *Psikologi belajar: Materi dasar pendidikan program bimbingan dan konseling di perguruan tinggi*. Jakarta: Ditjen Pendidikan Tinggi.
- Viswanathan,M. 2005. *Measurement error and research design*. Newbury Park: SAGE Publication Inc
- Widayati, C.S.W. 2009. Komparasi beberapa Metode Estimasi Kesalahan Pengukuran. *Jurnal Penelitian dan Evaluasi Pendidikan*, 13 (2), 1-16