



Sharing Pengetahuan Analisis Butir Tes dan Praktiknya Menggunakan Program R

Hari Purnomo Susanto¹⁾, Koirul Qudsiyah²⁾, Nely Indra Meifiani³⁾, Dwi Cahyani Nur Apriyani⁴⁾, Taufik Hidayat⁵⁾, Mega Isvandiana Purnamasari⁶⁾

^{1,2,3,4,5,6}STKIP PGRI Pacitan

e-mail: haripsusanto@stkippacitan.ac.id¹

Received: 1 Oktober 2022

Accepted: 10 Oktober 2022

Final proof: 30 Oktober 2022

Abstrak

Kalibrasi butir tes menggunakan konsep IRT masih dikuasai oleh beberapa pengajar di STKIP PGRI Pacitan. Konsep IRT memiliki banyak kelebihan jika dibandingkan dengan CTT. Tujuan dari kegiatan pengabdian masyarakat ini yaitu, 1) Melakukan sharing pengetahuan CTT dan IRT, 2) Pelatihan kalibrasi butir tes dengan menggunakan CTT dan IRT menggunakan program R untuk instrumen unidimensi. Sasaran dari kegiatan ini yaitu Pengajar Pan program Pan matematika, dan perwakilan dari setiap prodi di STKIP PGRI Pacitan. Metode yang digunakan dalam kegiatan ini yaitu (1) Metode Ceramah untuk menyampaikan materi pengetahuan kalibrasi dengan CTT dan IRT. 2) Metode diskusi dilakukan pada setiap pertemuan setelah materi pengetahuan dan pelatihan dilaksanakan. (3) Metode Pelatihan, dilakukan dengan mempraktikkan secara langsung Langkah-langkah dalam melakukan kalibrasi butir tes dengan CTT dan pada khususnya dengan IRT. Hasil dari kegiatan ini yaitu 1) 100 % peserta merefresh pengetahuannya tentang konsep CTT dan memiliki pengetahuan baru terkait kalibrasi butir dengan IRT. 2) 100% peserta mampu mengaplikasikan kalibrasi butir dengan IRT menggunakan program R. 3) peserta dapat membedakan CTT dan IRT, serta kapan harus menggunakan salah satu dari konsep tersebut.

Kata Kunci: Kalibrasi Butir tes, IRT, CTT, Program R, Rstudio, dan irtGUI

Abstract

Some teachers at STKIP PGRI Pacitan still master the calibration of test items using the IRT concept. The IRT concept has many advantages when compared to CTT. The objectives of this community service activity are, 1) Sharing CTT and IRT knowledge, 2) Test item calibration training using CTT and IRT using the R program for unidimensional instruments. The targets of this activity are the Education Teachers for the Mathematics Education program, and representatives from each study program at STKIP PGRI Pacitan. The methods used in this activity are (1) Lecture Method to convey calibration knowledge material with CTT and IRT. 2) The discussion method is carried out at each meeting after the knowledge and training materials are implemented. (3) Training Method, carried out by practicing directly the steps in calibrating test items with CTT and in particular with IRT. The results of this activity were 1)

100% of participants refreshed their knowledge about the CTT concept and had new knowledge related to item calibration with IRT. 2) 100% of participants are able to apply item calibration with IRT using the R program. 3) Participants can differentiate between CTT and IRT, and when to use one of these concepts.

Key Word: *item Calibration, IRT, CTT, Program R, Rstudio, dan irtGUI*

PENDAHULUAN

Pengembang tes pada dasarnya memperhatikan kualitas item tes dan bagaimana peserta ujian menanggapi saat membuat tes. Seorang psikometri umumnya menggunakan psikometri untuk menentukan validitas dan reliabilitas. Teori psikometri menawarkan dua pendekatan dalam menganalisis data tes: teori tes klasik (CTT) dan teori respons item (IRT). Kedua teori memungkinkan untuk memprediksi hasil tes psikologi dengan mengidentifikasi parameter kesulitan item dan kemampuan peserta tes. Keduanya berkepentingan untuk meningkatkan reliabilitas dan validitas tes psikologi. Kedua pendekatan ini memberikan ukuran validitas dan reliabilitas. Ada beberapa masalah yang diidentifikasi dalam teori tes klasik yang berkaitan dengan kalibrasi kesulitan item, ketergantungan sampel dari ukuran koefisien, dan perkiraan kesalahan pengukuran yang pada gilirannya ditangani oleh teori respon item. Tujuan artikel ini untuk menunjukkan keuntungan dan kerugian menggunakan kedua pendekatan dalam kalibrasi butir.

Classical Test Theory (CTT) dianggap sebagai “teori skor sejati”. Teori dimulai dari asumsi bahwa efek sistematis antara tanggapan peserta ujian hanya disebabkan oleh variasi dalam kemampuan minat. Semua sumber variasi potensial lainnya yang ada dalam bahan pengujian seperti: kondisi eksternal atau kondisi internal peserta ujian diasumsikan konstan melalui standarisasi yang ketat atau memiliki efek yang nonsistematis atau acak secara alami (Van derLinden & Hambleton, 2004). Model sentral dari teori tes klasik adalah tes yang diamati skor (TO) terdiri dari skor benar (T) dan skor kesalahan E dan lebih umum dituliskan $TO = T + E$.

Hambleton et al., (1991) menyebutkan beberapa kelemahan CTT, diantaranya (1) karakteristik butir tes teralalu bergantung pada karakteristik peserta tes; (2) kemampuan peserta tes tergantung pada butir tes; (3) *standar error* estimasi skor yang dihasilkan hanya untuk tes, sedangkan standar error estimasi skor untuk tiap peserta tes tidak dapat ditentukan. (4) hanya mampu menyajikan informasi sebatas benar atau salah saja tidak memperhatikan pola jawaban peserta tes; dan (5) asumsi tes susah dipenuhi.

Item Response Theory (IRT) dengan teori sifat laten atau sebagai teori skor sejati yang kuat atau teori tes mental modern. IRT merupakan teori yang lebih baru dan memiliki asumsi yang lebih kuat dibandingkan dengan CTT. Pendekatan pengujian berdasarkan analisis item ini mempertimbangkan peluang untuk mendapatkan item tertentu dengan benar atau salah. Dalam pendekatan ini, setiap item pada sebuah tes memiliki kurva karakteristik itemnya sendiri yang menggambarkan probabilitas untuk mendapatkan setiap item tertentu benar atau salah berdasarkan kemampuan peserta tes (Kaplan & Saccuzzo, 1997). Model Rasch sebagai contoh IRT cocok untuk memodelkan respons dikotomis dan memodelkan probabilitas respons yang benar individu pada item dikotomis. Kurva karakteristik item, fungsi kemampuan, membentuk batas antara area probabilitas menjawab item salah dan menjawab item dengan benar. Model 120 satu parameter ini mengasumsikan bahwa diskriminasi semua item diasumsikan sama dengan satu (Maier, 2001).

Ada beberapa perbedaan antara CTT dan IRT. Pertama, pada CTT kemampuan peserta tes bergantung pada instrument tes. Sebaliknya, Pada IRT kemampuan peserta

tes tidak bergantung pada instrument tes, sehingga bersifat independent. seseorang sepenuhnya bergantung pada serangkaian item tertentu karena skor sebenarnya ditentukan dalam kaitannya dengan tes atau skala tertentu. Kedua, pada CTT standar error hanya berlaku secara grup, sedangkan dalam IRT standar error untuk setiap item dan kemampuan dapat ditentukan (Allen & Yen, 1979)(Yen, 2002). Ketiga, Pada CTT parameter butir tes yang dihasilkan sangat bergantung pada kemampuan peserta tes. Sedangkan, pada IRT, Parameter butir tidak bergantung pada karakteristik kemampuan, begitu juga sebaliknya parameter kemampuan tidak bergantung pada parameter butir test. selanjutnya pada IRT karakteristik butir tes dapat digunakan untuk mengestimasi kemampuan individu(Shankness & DeAngelo, 2011).

Berdasarkan Informasi di atas, bukan berarti IRT tidak memiliki kelemahan. Beberapa kelemahan dari IRT adalah, pertama. Membutuhkan data sampel yang lebih banyak untuk mendapatkan karakteristik butir tesnya, dimana CTT cukup dengan sampel yang lebih kecil. Sampel minimum yang dapat digunakan untuk model IRT dikotomis unidimensi yaitu 150 (Şahin & Anil, 2017), atau 200 (Suwanto et al., 2019). Pada model politomis *Generalized Partial Credit Model* (GPCM) 200 sampel dapat digunakan (Huang, 2016), dan Pada *Graded Response Model* (GRM) membutuhkan 500 sampel (Reise & Yu, 1990). Kelemahan ini, sampai saat ini masih belum bisa diatasi oleh para ahli. Kedua, perhitungan yang kompleks juga menjadi kelemahan dari IRT, jika dibandingkan dengan CTT. Tetapi seiring dengan perkembangan komputasi kelemahan IRT ini dapat diatasi dan juga banyak *software* yang dapat digunakan baik berbayar atau gratis. Beberapa program yang gratis untuk digunakan yaitu R, dan Python.

Berdasarkan respon peserta tes pada instrumen yang digunakan atau dikembangkan. Baik pada CTT atau IRT, maka instrument dibagi menjadi dua yaitu dikotomis dan politomis. Pada kalibrasi instrument dikotomis menggunakan CTT, sudah pasti dapat dihasilkan tiga parameter butir tes, yaitu daya pembeda (Diskriminan), Tingkat Kesulitan dan tebakan. Ketiga parameter ini dapat juga dijumpai pada IRT dengan instrument dikotomis untuk model tiga parameter logistic (3PL). selain 3PL, pada IRT juga terdapat model 1PL dan 2PL (DeMars, 2010; Hambleton et al., 1991). Model 1PL hanya memiliki satu parameter butir yaitu tingkat kesulitan. Selanjutnya model 2PL memiliki dua parameter yaitu diskriminan dan tingkat kesulitan (Retnawati, 2014). Pada IRT dikotomis penentuan instrument cocok dengan model 1PL, 2PL atau 3PL didasarkan pada perbandingan model saat kita melakukan kalibrasi. Tetapi, jika kalibrasi digunakan untuk tujuan tertentu, maka langsung dapat dipilih salah satu model yang diinginkan sesuai tujuan.

Selanjutnya pada instrument politomis, baik CTT atau IRT juga dapat menunjukkan kemampuannya dalam proses kalibrasi. Tetapi lagi-lagi IRT menunjukkan kelebihan dalam hal ini. IRT memiliki banyak model yang dapat digunakan pada instrument politomis. Model respon butir yang dapat digunakan untuk menganalisis butir politomis diantaranya *Nominal Respons Model* (NRM), *Rating Scale Model* (RSM), *Partial Credit Model* (PCM), *Graded Respon Model* (GRM) dan *Generalized Partial Credit Model* (GPCM) (Vander Linden & Hambleton, 1997).

Sesuai hasil Observasi tim, pada STKIP PGRI Pacitan diperoleh bahwa semua pengajar pengembangan instrument masih menggunakan CTT sebagai alat untuk melakukan kalibrasi instrument. Update pengetahuan analisis butir tes khususnya untuk konsep IRT perlu dilakukan dengan beberapa pertimbangan, pertama sebagai alternatif dari CTT, jika kondisi memungkinkan. Kedua, sangat efektif digunakan untuk sampel yang lebih luas karena instrument tidak dipengaruhi oleh kemampuan peserta atau sebaliknya, sehingga cocok untuk penelitian dengan pendekatan kuantitatif. Ketiga, hasil kalibrasi dapat memberikan gambaran terkait individu beserta informasi satandar error tiap individu, dimana pada CTT hanya bisa diakses pada standar error grup saja. Keempat, Parameter butir tes yang diperoleh dari proses Kalibrasi dapat

digunakan untuk *scaling* atau estimasi kemampuan peserta tes. Beberapa pertimbangan ini bukan berarti kita harus menggunakan IRT, tetapi IRT memang memiliki banyak kelebihan jika dibandingkan dengan CTT.

Software-software analisis IRT yang dapat digunakan dengan mudah tanpa harus melakukan koding kebanyakan berbayar. Sehingga tidak menarik untuk digunakan, karena dalam proses publikasi internasional pada umumnya akan menanyakan lisensi dari *software* yang digunakan. Lisensi yang tidak berbayar mungkin akan lebih aman jika analisis yang digunakan untuk publikasi. *Software-software* yang bebas dari lisensi yaitu program R dan Python. Program R kelebihan jika dibandingkan dengan python dalam hal analisis data, walaupun diakhir-akhir ini banyak pengembang python mengarah pada bidang komputasi dan analisis data. Selanjutnya python memiliki keunggulan dalam pengembangan *software* yang dapat berdiri sendiri, sedangkan pada R untuk bisa menjadi *software* yang berdiri sendiri membutuhkan package compiler yang tidak mudah untuk digunakan. Berdasarkan informasi tentang *software* diatas, maka *software* yang digunakan kegiatan sharing ini yaitu Program R. selanjutnya *software* yang digunakan untuk kaliberasi menggunakan CTT dapat digunakan iteman atau program R.

Berdasarkan informasi-informasi diatas, maka tujuan tim yaitu, 1) sharing analisis butir tes menggunakan konsep IRT dan CTT. 2) Latihan analisis menggunakan program R untuk kaliberasi instrument dengan IRT.

METODE

Kegiatan ini dilakukan dengan menggunakan metode Ceramah, Diskusi, dan pelatihan. Metode ceramah digunakan untuk menyampaikan materi pengetahuan tentang CTT dan IRT. Metode diskusi digunakan untuk menyamakan persepsi pemahaman tentang CTT dan apa yang belum dipahami pada IRT. Terakhir metode pelatihan lebih ditekankan pada Langkah-langkah proses kalibrasi butir tes menggunakan konsep IRT dengan bantuan Program R.

Kegiatan dilaksanakan dalam empat pertemuan sebagai berikut

- a. Pertemuan pertama
Pada pertemuan ini dilakukan menyamakan persepsi tentang pemahaman analisis butir tes menggunakan CTT dan praktik analisis CTT menggunakan program Anates.
- b. Pertemuan kedua
Pada pertemuan ini dijelaskan konsep analisis butir tes menggunakan konsep IRT untuk instrument dikotomus unidimensi.
- c. Pertemuan ketiga
Pada pertemuan ini dilakukan panduan untuk mengisntal program R dan praktik analisis butir tes dengan konsep IRT menggunakan program R
- d. Pertemuan keempat
Pada pertemuan ini dilakukan praktik analisis butir tes dengan konsep IRT menggunakan program R

Keempat pertemuan diatas dilaksanakan sesuai materi, pemateri dan waktu pelaksanaan seperti pada table 1 di bawah.

Tabel 1. Jadwal kegiatan sharing analisis butir tes

No	Materi	Tanggal pelaksanaan	Pemateri	Tempat
1.	Kalibrasi item menggunakan konsep CTT	4 Maret 2022	Khoirul Qudsyah	STKIP PGRI Pacitan
2.	Analisis CTT menggunakan <i>software</i>			

No	Materi	Tanggal pelaksanaan	Pemateri	Tempat
	Anates			
3.	Kalibrasi item menggunakan konsep IRT	5 Mei 2022		STKIP PGRI Pacitan
4.	Pengenalan program R dan Analisis IRT menggunakan <i>software</i> R Studio	7 Juli 2022	Hari Purnomo Susanto	Zoom
5	Analisis IRT menggunakan <i>software</i> R Studio	8 Juli 2022		STKIP PGRI Pacitan

HASIL DAN PEMBAHASAN

1. Peremuan Pertama

Materi CTT disampaikan pada kegiatan ini, hanya untuk menyamakan persepsi diantara pengajar-pengajar dari setiap prodi. Sebelum menyamakan persepsi diantara peserta, pemateri menyampaikan konsep-konsep utama dari kalibrasi instrument menggunakan CTT. Setelah materi CTT, pemateri mengajak peserta untuk diskusi untuk menyamakan persepsi. Penyamaan persepsi dilakukan karena substansi instrumen yang dikembangkan oleh setiap peserta berbeda-beda. Beberapa hal yang menjadi bahan diskusi dapat dilihat pada tabel 2. Setelah diskusi, pemateri mengajak semua peserta untuk melakukan kalibrasi menggunakan konsep CTT dengan *software* anates.



Gambar 1. Pelaksanaan Kegiatan

Tabel 2. Hasil diskusi CTT

No	Bahan diskusi	Hasil diskusi
1.	Parameter butir tes a. Parameter kesulitan b. Parameter daya beda c. Parameter	Parameter-parameter butir ini hanya memberikan informasi yang digunakan untuk mengelompokkan butir tes, tidak digunakan untuk scaling data baru dari hasil tes selanjutnya.
2.	True skor	$T = T + E$, True skor untuk setiap individu tidak dapat ditentukan, karena eror pengukuran untuk setiap peserta tidak dapat ditentukan. Sehingga rumus tersebut hanya sebatas konsep yang dalam kenyataannya sulit untuk dibuktikan dengan data empiris.
3.	Standar Error	Standar Error yang dapat diakses hanya standar eror dari tes secara keseluruhan, dan error untuk setiap butir tes tidak dapat ditentukan. Selanjutnya penentuan kemampuan individu juga tidak dapat ditentukan Errornya.
4.	Kapan Menggunakan CTT?	CTT digunakan ketika tes yang dilakukan memiliki lingkup yang kecil. Misalnya tes pada suatu sekolah memberikan informasi yang sangat sesuai terhadap kemampuan siswanya. Tes tersebut belum tentu akan memberikan informasi yang sama ketika diberikan disekolah yang lain. Hal ini dikarenakan karakteristik tes tergantung pada karakteristik peserta tes.
5	Aplikasi yang bisa digunakan	Exel, Itean, Anates

2. Pertemuan kedua

Pemateri menyampaikan konsep IRT dan Langkah-langkah kalibrasi dengan IRT selama 30 menit. Penyampaian materi IRT dilakukan dengan dua bagian. Bagian pertama, pemateri menyampaikan logika berpikiri dari terbentuknya konsep IRT melalui model rasch, dilanjutkan menjelaskan jenis-jenis model IRT, dan diakhiri dengan menjelaskan kelebihan dan kekurangan dari konsep IRT. Selanjutnya, pemateri mengajak peserta untuk berdiskusi terkait dengan konsep IRT dan perbedaannya dengan konsep CTT. Bagian kedua, pemateri menjelaskan Langkah-langkah untuk melakukan kalibrasi butir tes beserta uji asumsi yang harus terpenuhi. Selanjutnya, pemateri mengajak peserta untuk berdiskusi terkait Langkah-langkah analisis kalibrasi dengan IRT.

Tabel 3. Hasil Diskusi IRT

No	Pertanyaan saat diskusi	Hasil diskusi
1.	Banyak dan jenis model IRT?	1PL, 2PL, dan 3PL GRM, PCM, GPCM, dan MGRM, dll.
2.	Bagaimana menentukan model IRT yang sesuai dengan instrumen?	1. Menentukan jenis instrument kita termasuk pada dikotomous atau politomus. 2. Menentukan kecocokan model dengan bantuan <i>software</i> . atau 3. Ditentukan sejak awal sesuai dengan tujuan tertentu, misal model di sesuaikan dengan model A, maka data yang diperoleh harus dikalibrasi dengan model A secara langsung.
3.	Banyak data yang dibutuhkan untuk kalibrasi?	Sesuai hasil penelitian, minimal banyak data yang dibutuhkan untuk masing-masing model berbeda-beda. 150 data untuk model 1 PL, 250 sampai 500 untuk 2 PL. dan lebih dari 500 untuk 3 PL. walaupun secara teknis dengan data kurang dari 150 tetap bisa digunakan untuk analisis.
4.	Parameter butir tes pada CTT hanya untuk karakteristik butir saja, selanjutnya bagaimana parameter butir dari IRT.	Parameter yang dihasilkan dari kalibrasi butir tes dengan IRT, tidak hanya digunakan untuk menentukan karakteristik butir tes, tetapi dapat digunakan untuk mengestimasi parameter kemampuan dari peserta tes.
5.	Perbedaan parameter ability pada IRT dan CTT?	Parameter kemampuan dari setiap peserta dapat ditentukan beserta standar erornya. Dimana pada CTT standar eror untuk setiap peserta tes tidak dapat ditentukan.
6.	Fungsi Informasi tes?	Memberikan informasi dari keseluruhan instrument cocok untuk diujikan pada peserta dengan kemampuan tertentu.
7.	Fungsi Informasi butir?	Memberikan informasi dari setiap butir tes cocok untuk diujikan pada peserta dengan kemampuan tertentu.
8.	Standar Error	Memberikan penjelasan kemungkinan error yang dapat terjadi jika tes digunakan.
9.	Kegunaan hasil kalibrasi IRT	1. Dapat digunakan untuk CAT, 2. Dapat digunakan untuk diagnosis otomatis 3. Dapat digunakan untuk tes untuk sampel yang lebih luas. 4. Pada instrument dikotomous dapat digunakan untuk melihat peserta yang mengerjakan sesuai kemampuannya atau diduga asal memilih jawaban. 5. Hasil kalibrasi butir dapat dikumpulkan ke dalam sebuah bank butir, dimana butir tes

No	Pertanyaan saat diskusi	Hasil diskusi
		dapat dipilih dengan kondisi tertentu untuk tujuan seleksi, diagnosis, dan lain-lain
10.	Bagaimana mengatasi jika uji unidimensi tidak terpenuhi?	Jika pembuktian asumsi unidimensional ini tidak terbukti, maka instrument kita harus dianalisis menggunakan IRT multidimensional.
11.	Bagaimana mengatasi jika uji independensi local tidak terpenuhi?	Jika terjadi kasus independensi local dapat diatasi dengan menghapus item yang menyebabkan terjadinya independensi lokal.
12.	Kapan IRT digunakan?	Jika tes yang digunakan mencakup wilayah sekolah yang lebih luas, dan memiliki standar Bersama. Karakteristik tes tidak dipengaruhi oleh karakteristik peserta tes, atau sebaliknya.
13.	Aplikasi yang dapat digunakan	Program R, bilog,

3. Pertemuan ketiga

Pada pertemuan ketiga, pemateri mengajak peserta untuk menginstal program R dan praktik beberapa kasus kalibrasi butir tes dengan IRT menggunakan program R. pertemuan ini dilakukan dengan zoom. Penggunaan program R untuk pemula yang belum pernah menggunakan minimal program Matlab, Maple, dll atau Bahasa pemrograman dasar akan mengalami kesulitan. Penggunaan aplikasi tambahan visual studio akan memberikan kemudahan bagi siapapun yang akan menggunakan program R. lebih lanjut pemateri menggunakan package GUI yang tanpa melakukan coding, sehingga lebih memudahkan peserta dalam analisis IRT unidimensional yaitu menggunakan package irtGUI.

Selanjutnya, pemateri memberikan contoh kasus yang memenuhi setiap Langkah pada analisis kalibrasi butir dengan IRT. Pemateri menjelaskan Langkah-langkah analisis dan interpretasi hasil dengan sangat jelas, dan tidak banyak pertanyaan dari peserta.

4. Pertemuan keempat

Pada pertemuan keempat, pemateri memberikan beberapa kasus kalibrasi butir dengan IRT dalam beberapa kondisi. Pertama jika ada uji asumsi yang tidak terpenuhi dan bagaimana mengatasinya. Kedua, memilih model yang cocok berdasarkan pembuktian analisis dan berdasarkan tujuan kalibrasi yang diinginkan.

Kegiatan sharing ini memberikan pengetahuan baru pada peserta, karena sebelumnya peserta hanya mengenal CTT. Pengetahuan tentang IRT akan memberikan suasana baru dalam melakukan kalibrasi instrumen, walaupun tidak mudah, karena untuk menguasai IRT tidak bisa hanya dengan beberapa pertemuan sharing. Pada praktik yang sebenarnya praktisi dapat menjumpai kasus-kasus yang mungkin tidak tersampaikan selama kegiatan sharing berlangsung. Selanjutnya, Agar pengetahuan yang dimiliki oleh peserta semakin mendalam dan bertransformasi menjadi keterampilan, maka kegiatan sharing dikemas menjadi satu dengan praktik. Setiap peserta diberi data dan diberikan kesempatan untuk mempraktikkan kalibrasi menggunakan konsep CTT dan IRT. Pada kesempatan ini, *software* yang digunakan oleh pemateri yaitu Anates untuk CTT, dan Program R untuk IRT.

SIMPULAN

Berdasarkan deskripsi pelaksanaan kegiatan dapat disimpulkan (1) 100 % peserta merefresh pengetahuannya tentang konsep CTT dan memiliki pengetahuan baru terkait kalibrasi butir dengan IRT. (2) 100% peserta mampu mengaplikasikan kalibrasi butir dengan IRT menggunakan program R. (3) peserta dapat membedakan CTT dan IRT, serta kapan harus menggunakan salah satu dari konsep tersebut.

UCAPAN TERIMA KASIH

Terimakasih kepada seluruh peserta yang berkenan dalam mengikuti sharing kegiatan pengabdian masyarakat ini.

DAFTAR PUSTAKA

- Allen, M. J., & Yen, W. M. (1979). Introduction to measurement theory, Monterey, CA: Brooks/Cole, 1979. *Google Scholar*.
- DeMars, C. (2010). Item Response Theory. In *Item Response Theory*. <https://doi.org/10.1093/acprof:oso/9780195377033.001.0001>
- Hambleton, R., Swaminathan, H., & Rogers, H. J. (1991). *fundamental of item response theory*. SAGE.
- Huang, H.-Y. (2016). Mixture Random-Effect IRT Models for Controlling Extreme Response Style on Rating Scales. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.01706>
- Reise, S. P., & Yu, J. (1990). Parameter Recovery in the Graded Response Model Using MULTILOG. *Journal of Educational Measurement*, 27(2). <https://doi.org/10.1111/j.1745-3984.1990.tb00738.x>
- Retnawati, H. (2014). *Teori Respons Butir dan Penerapannya*.
- Şahin, A., & Anıl, D. (2017). The Effects of Test Length and Sample Size on Item Parameters in Item Response Theory. *Educational Sciences: Theory & Practice*. <https://doi.org/10.12738/estp.2017.1.0270>
- Sharkness, J., & DeAngelo, L. (2011). Measuring Student Involvement: A Comparison of Classical Test Theory and Item Response Theory in the Construction of Scales from Student Surveys. *Research in Higher Education*, 52(5), 480–507. <https://doi.org/10.1007/s11162-010-9202-3>
- Suwarto, Widoyoko2, E. P., & Setiawan, B. (2019). The Effects of Sample Size and Logistic Models on Item Parameter Estimation. *Proceedings of the 2nd International Conference on Education*, 323–330.
- Vander Linden, W. J., & Hambleton, R. K. (1997). Handbook of Modern Item Response Theory. *Journal of the American Statistical Association*, 92(439). <https://doi.org/10.2307/2965612>