

Pengungkapan Sebaran Kesukaran Soal Tes Berbasis HOTS melalui Pemodelan Rasch

Ade Yulianto¹, Ari Widodo²

^{1,2} Program Studi Magister Pendidikan Dasar, Sekolah Pascasarjana Universitas Pendidikan Indonesia, Bandung.

*Corresponding author: adeyulianto@upi.edu¹, widodo@upi.edu²

Received 8 November 2020; Revised 16 November 2020; Accepted 10 December 2020

Published 23 December 2020

Abstract

Many researchers or teachers need more and more complete information about the evaluation of the implementation of learning or the ability measurement tools used to find out how much learning outcomes are shown by their students. In this article, it is shown how to analyze questions based on the level of difficulty and suitability of test item items; especially on HOTS (Higher-Order Thinking Skill) based test questions which were developed based on a cognitive hierarchy adopted from Bloom's taxonomy (C4, C5, & C6). The analysis process was carried out based on Sumintono's (2015) explanation of the combination of standard deviation (SD) values and logit average values (Mean). Then perform the criteria for the outfit mean square (MNSQ) value, the Z-standard outfit value (ZSTD) and the point measure correlation value (Pt. Measure Corr) (Boone et al., 2014). The analysis technique was carried out through Rasch modeling assisted by the Winsteps 3.75 application. As for the grouping of difficulty levels according to Sumintono (2015), namely 1) difficult question category (logit value is greater + 1SD); 2) difficult question category (value 0.0 logit + 1 SD); 3) easy question category (value 0.0 logit - 1 SD); and 4) the category of questions is very easy (value less than -SD), as well as for the criteria used to measure the suitability of item items using, 1) the value of $0.5 < MNSQ < 1.5$; 2) value $-2.0 < ZSTD < +2.0$; and 3) the value of $0.4 < PT\text{-Measure Corr} < 0.85$, (Boone et al., 2014). The results of the analysis show that there are variations in the level of difficulty and suitability of HOTS-based test items. The item analyzed had an acceptable level of suitability and was feasible to maintain because all item items met these three criteria. Thus, the collection of HOTS-based test questions is in a good category because it can identify students' various abilities in higher-order thinking with varying levels of difficulty.

Keywords: Analysis of Item Difficulty and Suitability; HOTS-Based Test Questions; Rasch Modeling.

Abstrak

Banyak peneliti atau guru yang perlu informasi lebih banyak dan lengkap tentang evaluasi pelaksanaan pembelajaran atau alat pengukuran kemampuan yang digunakan untuk mengetahui seberapa besar hasil belajar yang ditunjukkan oleh siswanya. Pada artikel ini, ditunjukkan bagaimana menganalisis soal berdasarkan pengelompokan tingkat kesukaran dan kesesuaian item soal tes; khususnya pada soal tes berbasis HOTS (Higher-Order Thinking Skill) yang dikembangkan berdasarkan hirarki kognitif yang diadopsi dari taksonomi Bloom (C4, C5, & C6). Proses analisis dilakukan berdasar pada penjelasan Sumintono (2015) tentang kombinasi nilai standar deviasi (SD) dan nilai rata-rata logit (Mean). Selanjutnya melakukan kriteria nilai outfit mean square (MNSQ), nilai outfit Z-standard (ZSTD) serta nilai point measure correlation (Pt. Measure Corr) (Boone et al., 2014). Teknik analisis dilakukan melalui pemodelan Rasch berbantuan aplikasi Winsteps 3.75. Adapun pengelompokan tingkat kesukaran menurut Sumintono (2015) yakni 1) kategori soal sulit (nilai logit lebih besar +1SD); 2) kategori soal sukar (nilai 0,0 logit + 1 SD); 3) kategori soal mudah (nilai 0,0 logit - 1 SD); dan 4) kategori soal sangat mudah (nilai lebih kecil dari -SD), serta untuk kriteria yang digunakan untuk mengukur kesesuaian item soal menggunakan, 1) nilai $0,5 < MNSQ < 1,5$; 2) nilai $-2,0 < ZSTD < +2,0$; dan 3) nilai $0,4 < PT\text{-Measure Corr} < 0,85$, (Boone et al., 2014). Hasil analisis menunjukkan bahwa terdapat keberagaman tingkat kesukaran dan kesesuaian item soal tes berbasis HOTS. Item soal yang dianalisis memiliki tingkat kesesuaian yang dapat diterima dan layak dipertahankan karena seluruh item soal memenuhi ketiga kriteria tersebut. Dengan demikian, kumpulan soal tes berbasis HOTS berkategori baik karena dapat mengidentifikasi berbagai kemampuan siswa dalam berpikir tingkat tinggi dengan tingkat kesukaran yang beragam dan hal ini berimplikasi pada teknik guru dalam menyajikan soal tes tertulis yang berkualitas.

Kata Kunci: Analisis Kesukaran dan Kesesuaian Item; Soal Tes Berbasis HOTS; Pemodelan Rasch.

PENDAHULUAN

Pada proses pembelajaran tentunya memiliki tujuan belajar mengenai sejumlah

pengetahuan atau keterampilan yang harus dicapai oleh siswa. Dalam upaya mengetahui proses pencapaian ini dibutuhkan

alat ukur untuk mengetahui nilai, skor atau persentasi yang dicapai oleh siswa terkait dengan tujuan pembelajaran tersebut. Pada proses atau akhir pembelajaran perlu dilakukan pengukuran terhadap proses dan hasil belajar berupa angka-angka yang mencerminkan capaian proses dan hasil belajar. Menurut Mardapi (2017), pengukuran pada dasarnya adalah kegiatan penentuan angka terhadap suatu obyek secara sistematis. Pengukuran merupakan proses yang mendeskripsikan performance siswa dengan menggunakan suatu skala kuantitatif (sistem angka) sedemikian rupa sehingga sifat kualitatif dari performance siswa tersebut dinyatakan dengan angka-angka (Alwasillah et al, 1996). Dengan demikian, pengukuran dalam pendidikan berarti mengukur atribut atau karakteristik siswa (Safithry, 2019).

Pengukuran dalam bidang pendidikan erat kaitannya dengan tes. Hal ini dikarenakan salah satu cara yang sering dipakai untuk mengukur hasil yang telah dicapai siswa adalah dengan tes (Tillar, 2018). Dalam proses pengukuran, tentu guru harus menggunakan alat ukur baik tes atau non-tes (Sadikin, 2018). Menurut Zainul (2001), tes didefinisikan sebagai pertanyaan atau tugas atau seperangkat tugas yang direncanakan untuk memperoleh informasi tentang suatu atribut pendidikan. Tes yang digunakan disesuaikan dengan mata pelajaran atau bidang ilmu yang akan dijadikan sumber tes. Adapun soal tes yang dikembangkan untuk menggali kemampuan siswa berpikir tingkat tinggi. Kemampuan berpikir tingkat tinggi perlu dikembangkan dalam pendidikan di sekolah dasar sebagai upaya untuk meningkatkan kemampuan siswa berpikir nalar dalam menjawab pertanyaan-pertanyaan yang lebih rumit dan atau memecahkan suatu kasus masalah yang lebih rumit. Heong (2011) menjelaskan bahwa kemampuan berpikir tingkat tinggi menghendaki seseorang untuk menerapkan informasi baru atau pengetahuan sebelumnya dan memanipulasi informasi untuk menjangkau kemungkinan jawaban dalam situasi baru. Istilah kemampuan berpikir tingkat tinggi atau *higher-order thinking skills* (HOTS) merupakan bagian dari klasifikasi

kemampuan berpikir yang disampaikan oleh Bloom. Bloom (dalam Anderson & Karthwohl, 2001) mengklasifikasi kemampuan berpikir ke dalam enam tingkatan, yaitu mengingat/C-1, memahami/C-2, menerapkan/C-3, menganalisis/C-4, mengevaluasi/ C-5 dan mencipta/C-6. Brookhart (2010) menyebutkan tingkatan berpikir yang termasuk ke dalam HOTS meliputi tiga kemampuan berpikir tertinggi pada taksonomi Bloom yaitu menganalisis, mengevaluasi dan mencipta.

Hasil yang diperoleh dari tes adalah data yang akan dianalisis berbantuan aplikasi Winsteps 3.75 melalui model skala penilaian Rasch. Model Rasch merupakan pendekatan pengukuran yang dipelajari dengan baik yang memodelkan hubungan antara kesukaran item, kemampuan orang, dan probabilitas respon yang diberikan (Andrich, 1981). Model Rasch yang menyediakan teknik analisis psikometrik dapat digunakan oleh guru untuk mengembangkan item tes serta alat penting yang dapat menyajikan informasi relevan terkait penilaian siswa untuk pembelajaran (Suminton, 2018). Analisis instrumen tes ini menggunakan model Rasch termasuk dalam teori pengukuran item respons. Pengukuran ini menjelaskan interaksi antara subjek dan item tes. Ini akan membuat pengukuran memiliki hasil yang lebih tepat dan obyektif (Sumintono & Widhiarso, 2014). Adapun menurut Brogde (1977), Model Rasch biasanya diterapkan pada pengukuran item dan subjek orang dalam konteks ini bahwa dibahas hubungannya dengan hukum penilaian perbandingan dan tambahan pengukuran gabungan lainnya. Menurut Masters (1982), Pemodelan Rasch dapat digunakan untuk berbagai format pengamatan antara lain model untuk analisis perhitungan, percobaan berulang dan skala penilaian. Selain itu, gambaran statistik kesesuaian Rasch dapat memberikan kerangka kerja yang berguna untuk menguji kebenaran respon seseorang, mengukur estimasi kemampuan respon seseorang dan mampu mendeteksi berbagai gangguan terhadap respon seseorang (Smith, 1986). Harus diperhatikan bahwa model Rasch ditulis sebagai model probabilitas respons individu terhadap suatu item dan

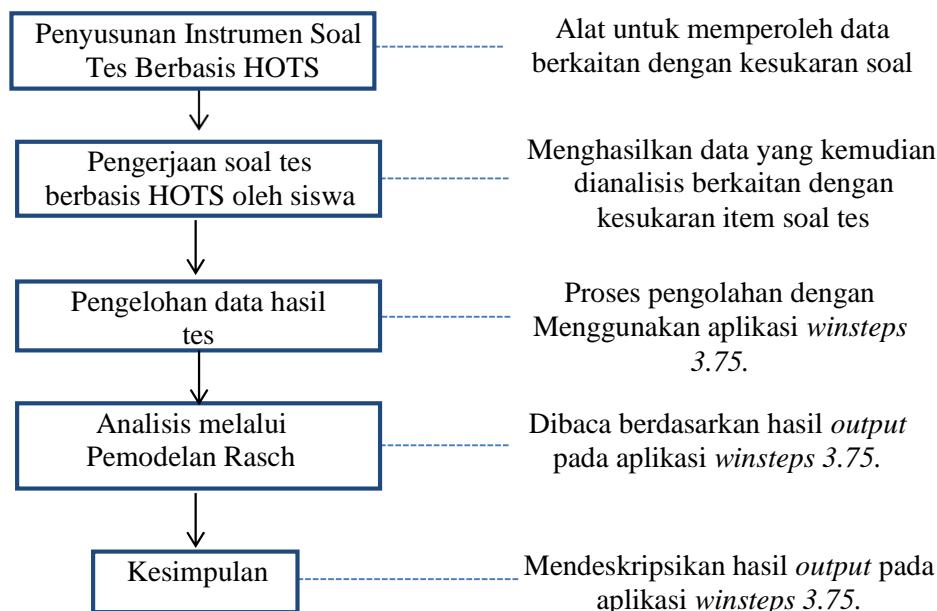
karenanya tidak secara eksplisit menjadi model respons itu sendiri (Brogden, 1977). Georg Rasch mengembangkan satu model analisis dari teori respon butir (atau Item Response Theory, IRT) pada tahun 1960-an, yang kemudian dipopulerkan oleh Ben Wright (Misbach & Sumintono, 2014). Dengan data mentah berupa data dikotomi (berupa benar dan salah) yang mengindikasikan kemampuan siswa, Rasch memformulasikan hal ini menjadi satu model yang menghubungkan antara siswa dan item (Sumintono & Widhiarso, 2015). Selain data dikotomi, model Rasch juga bisa melakukan analisis untuk data politomi seperti yang dikembangkan oleh Andrich, yang tetap berlandaskan pada dua teorema dasar, tingkat kemampuan seseorang dan tingkat kesukaran item.

Proses analisis melalui pemodelan Rasch pada penelitian terdahulu dilakukan dalam pengembangan instrumen karakter dan kecemasan di sekolah dasar, maka dalam penelitian ini menjadi rujukan dalam proses teknis pengolahan datanya (Nur et al., 2020; Karlimah et al., 2020) serta dalam mengukur kemampuan berpikir kritis siswa di sekolah dasar (Hamdu, et al., 2020). Model Rasch berasumsi bahwa kesukaran item adalah sifat yang dipengaruhi oleh jawaban responden, dan kemampuan seseorang adalah sifat yang

dipengaruhi oleh estimasi kesukaran item (Linacre, 1999). Keuntungan dari model Rasch bila dibandingkan dengan teori klasik adalah bahwa model ini dapat mengidentifikasi jawaban yang salah dari para ahli, mengidentifikasi penilaian yang tidak tepat, dan memprediksi data yang hilang berdasarkan pola respon sistematis (Pratama, 2020; Novinda et al., 2019). Secara spesifiknya penelitian ini bertujuan untuk mengungkap sebaran dan pengelompokan kesukaran soal tes dan kesesuaiannya (kelayakan) sehingga dapat memberikan gambaran kualitas soal yang telah dibuat dan evaluasi dalam pelaksanaan pembelajaran.

METODE PENELITIAN

Dalam proses pemerolehan data, di uji cobakan sebanyak 11 soal tes berbasis HOTS untuk dikerjakan kepada 20 siswa sekolah dasar yang sebelumnya telah melaksanakan pembelajaran pada tema 2 “Selalu Berhemat Energi”, Subtema “Manfaat Energi Alternatif” yang sesuai dengan materi pada soal yang dikembangkan. Setelah itu dilakukan analisis berkaitan kesukaran dan kesesuaiannya melalui pemodelan Rasch berbantuan aplikasi *winsteps 3.75*. Adapun tahapan proses analisis melalui pemodelan Rasch dengan adopsi tahapan yang dilakukan oleh Hamdu, et al. (2020), yakni:



Gambar 1
Tahapan Penelitian

HASIL DAN PEMBAHASAN

Analisis Tingkat Kesukaran Soal Tes Berbasis HOTS

Tabel 1 mengurai beberapa kolom yang dapat memberikan informasi mengenai tingkat kesukaran dari setiap butir soal tes berbasis HOTS. Pengelompokan tingkat kesukaran butir soal melalui kombinasi nilai

standar deviasi (SD) dan nilai rata-rata logit (Sumintono, 2015). Adapun pengelompokannya adalah kategori soal sulit dengan nilai logit lebih besar +1SD; kategori soal sukar dengan nilai 0,0 logit +1 SD; kategori soal mudah dengan nilai 0,0 logit -1 SD; kategori soal sangat mudah dengan nilai lebih kecil dari -SD.

Tabel 1
Hasil Output Item Measure

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PT-MEASURE CORR.	EXP.	EXACT OBS%	MATCH EXP%	Item
8	5	20	2.63	.86	1.20	.5	9.90	3.2	.57	.75	88.2	89.0	I8
1	8	20	1.02	.64	.61	-1.2	.42	-1.0	.82	.70	94.1	79.6	I1
2	9	20	.63	.61	1.25	.9	1.52	1.0	.58	.67	64.7	77.0	I2
4	10	20	.27	.59	1.09	.4	1.36	.9	.60	.65	70.6	75.8	I4
5	11	20	-.07	.58	1.01	.1	.90	-.1	.62	.62	82.4	74.6	I5
6	11	20	-.07	.58	1.37	1.3	1.29	.7	.51	.62	58.8	74.6	I6
11	11	20	-.07	.58	.65	-1.4	.52	-1.1	.74	.62	82.4	74.6	I11
3	13	20	-.74	.58	1.09	.4	1.08	.3	.52	.55	82.4	73.2	I3
7	13	20	-.74	.58	.83	-.6	.66	-.5	.62	.55	70.6	73.2	I7
9	15	20	-1.43	.61	.59	-1.7	.42	-.7	.62	.48	88.2	76.0	I9
10	15	20	-1.43	.61	.94	-.1	.70	-.2	.51	.48	76.5	76.0	I10
MEAN	11.0	20.0	.00	.62	.97	-.1	1.71	.2			78.1	76.7	
S.D.	2.9	.0	1.11	.08	.26	.9	2.62	1.2			10.4	4.3	

Berdasarkan Tabel 1 diatas, maka hasil HOTS dapat dikelompokkan sebagai berikut:

1. Kelompok butir soal sulit yakni soal no. I8.
2. Kelompok butir soal sukar yakni soal no. I1, I2, dan I4.
3. Kelompok butir soal mudah yakni soal no. I5, I6, I11, I3, dan I7.
4. Kelompok butir soal sangat mudah yakni soal no. I9 dan I10.

Tabel 2
Tingkat Kesukaran Berdasarkan Level Kognitif

No. Soal	Tingkat Kesukaran	Level Kognitif
1	Sukar	C4
2	Sukar	C4
3	Mudah	C4
4	Sukar	C4
5	Mudah	C6
6	Mudah	C5
7	Mudah	C4
8	Sulit	C4
9	Sangat Mudah	C5
10	Sangat Mudah	C5
11	Mudah	C4

Analisis Tingkat Kesesuaian Kesukaran Soal Tes Berbasis HOTS

Selanjutnya setelah dilakukan analisis tingkat kesukaran soal tes berbasis HOTS, lebih lanjut dilakukan analisis tingkat kesesuaian kesukaran butir soal (Item Fit) dapat dilihat dengan menggunakan tiga kriteria yakni nilai outfit means-square (Outfit MNSQ), Outfit Z-Standard (Outfit ZSTD), dan Point Measure Correlation (PT-Measure Corr) (Bonee et al., 2014; Bond & Fox, 2015;

Sumintono & Widhiarso, 2015). Kriteria yang digunakan untuk memeriksa kesesuaian butir soal yang tidak sesuai (outlier atau misfit), yakni:

1. Nilai outfit means-square (Outfit MNSQ) yang diterima: $0,5 < MNSQ < 1,5$.
2. Nilai Outfit Z-Standard (Outfit ZSTD) yang diterima: $-2,0 < ZSTD < +2,0$.
3. Nilai Point Measure Correlation (PT-Measure Corr): $0,4 < PT\text{-Measure Corr} < 0,85$.

Tabel 3
Hasil Output Item Fit Order

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		PT-MEASURE CORR.	EXP.	OBS%	MATCH EXP%	Item
					MNSQ	ZSTD	MNSQ	ZSTD					
8	5	20	2.63	.86	1.20	.5	9.90	3.2	A .57	.75	88.2	89.0	I8
2	9	20	.63	.61	1.25	.9	1.52	1.0	B .58	.67	64.7	77.0	I2
6	11	20	-.07	.58	1.37	1.3	1.29	.7	C .51	.62	58.8	74.6	I6
4	10	20	.27	.59	1.09	.4	1.36	.9	D .60	.65	70.6	75.8	I4
3	13	20	-.74	.58	1.09	.4	1.08	.3	E .52	.55	82.4	73.2	I3
5	11	20	-.07	.58	1.01	.1	.90	-.1	F .62	.62	82.4	74.6	I5
10	15	20	-1.43	.61	.94	-.1	.70	-.2	e .51	.48	76.5	76.0	I10
7	13	20	-.74	.58	.83	-.6	.66	-.5	d .62	.55	70.6	73.2	I7
11	11	20	-.07	.58	.65	-1.4	.52	-1.1	c .74	.62	82.4	74.6	I11
1	8	20	1.02	.64	.61	-1.2	.42	-1.0	b .82	.70	94.1	79.6	I1
9	15	20	-1.43	.61	.59	-1.7	.42	-.7	a .62	.48	88.2	76.0	I9
MEAN	11.0	20.0	.00	.62	.97	-.1	1.71	.2			78.1	76.7	
S.D.	2.9	.0	1.11	.08	.26	.9	2.62	1.2			10.4	4.3	

Dari Tabel 3 diatas, diperoleh bahwa butir soal I8 tidak memenuhi nilai MNSQ; butir soal I8 tidak memenuhi nilai ZSTD; dan seluruh butir soal memenuhi nilai PT-Measure Corr. Jika butir soal pada ketiga kriteria (MNSQ, ZSTD, dan Pt. Measure Corr) tidak terpenuhi dapat dipastikan soal kurang bagus sehingga perlu diperbaiki ataupun diganti (Bonee et al., 2014; Bond & Fox, 2015). Dari kutipan tersebut, maka untuk seluruh butir soal tes berbasis HOTS yang dianalisis memiliki tingkat kesesuaian yang dapat diterima dan layak dipertahankan karena tidak butir soal yang tidak memenuhi ketiga kriteria tersebut. Alasan ini dapat menunjukkan bahwa terdapat keberagaman tingkat kesukaran soal yang sesuai pada tabel 1. Kumpulan soal tes yang baik dapat mengidentifikasi berbagai kemampuan siswa dengan tingkat kesukaran yang beragam. Jika tingkat kesukaran dalam kumpulan soal tes tersebut tinggi maka dapat dipastikan bahwa siswa tidak bisa menjawab benar atau tidak

memahami soal yang diberikan. Namun begitu pula jika tingkat kesukaran dalam kumpulan soal tersebut rendah maka dapat dipastikan bahwa banyak siswa yang dapat menjawab benar atau mudah.

SIMPULAN

Berdasarkan hasil analisis dan pembahasan dapat disimpulkan bahwa sejumlah soal tes menunjukkan keberagaman tingkat kesukaran dengan tingkat kesesuaian yang dapat diterima. Adapun tingkat kesukaran berkaitan dengan materi bilangan pecahan yakni, 1) kelompok butir soal sulit mayoritas terdapat pada soal tes berbasis HOTS pada level kognitif C4; 2) kelompok butir soal sukar mayoritas terdapat pada soal tes berbasis HOTS pada level kognitif C4; 3) kelompok butir soal mudah mayoritas terdapat pada soal tes berbasis HOTS pada level kognitif yang bervariasi yakni C4, C5 dan C6; 4) kelompok butir soal sangat mudah mayoritas terdapat pada soal tes berbasis HOTS pada level kognitif C5.

Hasil analisis melalui pemodelan rasch ini lebih khusus untuk memberikan gambaran secara komprehensif pada pembelajaran pada tema 2 “Selalu Berhemat Energi”, Subtema “Manfaat Energi Alternatif” yang dilakukan pada subjek dan waktu tertentu. Hasil analisis pemodelan rasch ini dapat saja berbeda atau sama dengan memperhatikan kondisi dan situasi pembelajaran, seperti berkaitan dengan karakteristik siswa dan pelaksana pembelajaran di kelas atau sekolah tertentu. Namun proses analisis pemodelan rasch ini dapat digunakan para guru di sekolah untuk dapat melakukan identifikasi secara komprehensif mengenai proses pembelajaran dan pembuatan soal tes.

DAFTAR PUSTAKA

- Alwasillah, et al. (1996). *Glossary of educational Assessment Term*. Jakarta: Kemendikbud.
- Anderson & Krathwohl. (2001). *A Taxonomi for Learning, Teaching and Assessing: A Revision of Bloom’s Taxonomi of Educational Objectives*. New York: Addison Wesley Longman, Inc
- Andrich, D. (1981). Book Review : Probabilistic Models for Some Intelligence and Attainment Tests (expanded edition. *Applied Psychological Measurement*, 5(4), 545– 550. doi:10.1177/014662168100500413.
- Bond, T.G. & Fox, C.M. (2015). *Applying the Rasch Model, Fundamental Measurement in the Human Science* (3rd edition). New York: Routledge.
- Bonee, W.J., Staver, J.R., & Yale, M.S. (2014). *Rasch Analysis in the Human Science*. Dordrecht: Springer
- Brogden, H. E. (1977). The rasch model, the law of comparative judgment and additive conjoint measurement. *Psychometrika*, 42(4), 631–634. doi:10.1007/bf02295985.
- Brookhart, S M. (2010). *Higher-Order Thinking In Your Classroom*. ACDC, Alexandria, Virginia USA.
- Heong, Y. et.al. (2011). The level of marzano higher order thinking skills among technical education students. *International Journal of Social Science and Humanity*, 1(2), 121.
- Karlimah, K., Andriani, D., & Suryana, D. (2020). Development of Mathematical Anxiety Instruments with a Rasch Model Analysis. *The Open Psychology Journal*, 13(1).
- Linacre, J. M. (1999). Investigating Rating Scale Category Utility. *Journal of Outcome Measurement*. 3(2), 103-122.
- Mardapi, D. (2017). *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Yogyakarta: Parama Publishing
- Masters, G. N. (1982). A Rasch Model For Partial Credit Scoring. *Psychometrika*, 47(2), 149–174. doi:10.1007/bf02296272.
- Misbach, I.H. and Sumintono, B. (2014) Pengembangan dan validasi instrumen “persepsi siswa terhadap karakter moral guru” di Indonesia dengan Model Rasch. In: *Seminar Nasional Psikometri: "Pengembangan Instrumen Penilaian Karakter yang Valid"*, 24 May 2014, Solo, Indonesia.
- Novinda, M. R. R., Silitonga, H. T. M., & Hamdani, H. (2019). Pengembangan Tes Pilihan Ganda Menggunakan Model Rasch Materi Gerak Lurus Kelas X Pontianak. *Jurnal Pendidikan dan Pembelajaran Khatulistiwa*, 8(6).
- Nur, L., Nurani, L. A., Suryana, D., & Ahmad, A. (2020). Rasch model application on character development instrument for elementary school students. *International Journal of Learning, Teaching and Educational Research*, 19(3), 437-459.
- Pratama, D. (2020). Analisis Kualitas Tes Buatan Guru Melalui Pendekatan Item Response Theory (IRT) Model Rasch. *Tarbawy: Jurnal Pendidikan Islam*, 7(1), 61-70.
- Sadikin, A. (2018). *Evaluasi dan Proses Hasil belajar Biologi*. Jambi: Salim Media Indonesia.
- Saftihry, E. A. (2019). *Asesmen Teknik Tes dan Nontes*. Malang: CV. IRDH.

- Smith, R. M. (1986). Person Fit in the Rasch Model. *Educational and Psychological Measurement*, 46(2), 359–372. doi:10.1177/001316448604600210.
- Sumintono, B. (2018). Rasch Model Measurements as Tools in Assesment for Learning. In 1st International Conference on Education Innovation (ICEI 2017). Atlantis Press.
- Sumintono, B., & Widhiarso, W. (2015). Aplikasi Pemodelan RASCH pada Assessment Pendidikan. In TrimKom Publising Home.
- Tilaar, A.L.F. (2018). Asesmen Pembelajaran Matematika. Tondano: Unima Press.
- Zainul, A. (2001). Alternative assessment. Jakarta: Dirjen Dikti.