

Analisis Rasch Model: Mengukur Kualitas Butir Soal Tes *Tashrif Lughawi*

¹Salma Hayati, ²Sri Wahyuni Hidayanti, ³ Aulia Rizki

salma.hayati@ar-raniry.ac.id, 210202012@student.ar-raniry.ac.id,
aulia_rizki@ahsgs.uum.edu.my

^{1,2}Univeritas Islam Negeri Ar-Raniry Banda Aceh, Indonesia

³ Univeriti Utara Malaysia, Malaysia

Correspondence Address: salma.hayati@ar-raniry.ac.id

Abstract

This study aims to test the quality of the items of the Tashrif Lughawi using Rasch Model analysis. This study measured test items with item polarity tests, item and respondent mapping, individual and item reliability. This instrument contains 20 objective items test. This instrument test was tested on 51 students of the Madrasah Tsanawiyah level. The research data was analyzed using the Winsteps program for Rasch Model analysis. The result of Rasch's analysis based on item polarity shows that all test items are positive Point Measure Correlation (PMC) value. This result means that there is no conflict between the item and the constructed being measured. The results of the Mean Square Value show that 18 items have a value less than 1.5 which means that the resulting measurement is productive. 2 items -soal_10 (1.83) and soal_4 (1.64)- have a value greater than 1.5, so it is necessary to review this item for use decision-making. The results of the reliability analysis showed that the reliability index of the item was 0.91 and the reliability index of the respondents was 0.80. The results of this study show that the Tashrif Lughawi test instrument has high validity and reliability so that it can be used as a tool to measure the ability of Tashrif Lughawi students at the Madrasah Tsanawiyah level.

Keyword: Item Test, Tashrif Lughawi, Rasch Model Measurement

Abstrak

Penelitian ini bertujuan untuk menguji kualitas butir soal tes pengetahuan *Tashrif Lughawi* menggunakan analisis *Rasch Model*. Penelitian ini mengukur butir soal tes dengan uji *polarity item*, pemetaan item dan responden, reliabilitas individu dan item. Instrumen ini mengandung 20 soal tes berbentuk objektif. Soal tes ini diuji kepada 51 peserta didik jenjang Madrasah Tsanawiyah. Data penelitian dianalisis menggunakan program Winsteps untuk analisis *Rasch Model*. Analisis *Rasch* berdasarkan *polarity item* menunjukkan bahwa semua item soal mempunyai nilai *Point Measure Correlation* (PMC) yang positif. Hasil ini bermakna tidak terjadi

pertentangan antara item dengan konstruk yang diukur. Hasil Nilai Mean Square menunjukkan bahwa hanya 18 item soal memiliki nilai lebih kecil dari 1.5 yang bermakna pengukuran yang dihasilkan adalah produktif. 2 item soal -soal_10 (1.83) dan soal_4 (1.64)- memiliki nilai lebih besar dari 1.5 maka perlu dikaji kembali item ini untuk pengambilan keputusan penggunaan. Hasil analisis reliabilitas menunjukkan bahwa indeks reliabilitas item adalah 0.91 dan indeks reliabilitas responden adalah 0.80. Hasil penelitian ini menunjukkan bahwa instrumen soal tes pengetahuan Tashrif Lughawi memiliki validitas dan reliabilitas yang tinggi sehingga dapat digunakan sebagai alat ukur kemampuan *Tashrif Lughawi* peserta didik jenjang Madrasah Tsanawiyah.

Kata Kunci: Soal Tes, Tashrif Lughawi, Pengukuran Rasch Model

1. Introduction

Pendidikan bahasa Arab memiliki peranan penting dalam mengembangkan kemampuan berbahasa dan pemahaman agama di kalangan siswa (Fattah, 2019). Salah satu komponen kunci dalam pembelajaran bahasa Arab adalah *tashrif lughawi* atau morfologi bahasa Arab, yang memerlukan pemahaman mendalam tentang perubahan bentuk kata dan penggunaannya dalam konteks yang tepat (Samah et al., 2023). Evaluasi terhadap kemampuan *tashrif lughawi* peserta didik sering kali dilakukan melalui tes yang dirancang khusus untuk mengukur pemahaman dan keterampilan mereka dalam memahami *tashrif lughawi* (Putriyani & Qalbah, 2023).

Namun, kualitas butir soal dalam tes *tashrif lughawi* menjadi isu krusial yang harus diperhatikan oleh pendidik dan peneliti (Supandi & Farikhah, 2016). Butir soal yang tidak valid atau reliabel dapat memberikan hasil yang tidak akurat, yang pada akhirnya dapat mempengaruhi penilaian terhadap kemampuan peserta didik secara keseluruhan (Matondang, 2009). Oleh karena itu, diperlukan analisis mendalam untuk memastikan bahwa setiap butir soal memiliki kualitas yang baik dan mampu mengukur kemampuan siswa secara akurat dan konsisten.

Model Rasch, sebagai salah satu metode dalam teori respons butir (IRT), menawarkan pendekatan yang efektif untuk menganalisis kualitas butir soal. Model ini memungkinkan evaluasi terhadap kesesuaian butir soal dengan kemampuan peserta didik, serta mengidentifikasi butir soal yang mungkin tidak berfungsi dengan baik (Sumintono et al., 2014). Dalam konteks evaluasi pendidikan bahasa Arab, penerapan

Model Rasch ini dapat memberikan wawasan berharga tentang sejauh mana butir soal dalam tes *tashrif lughawi* memenuhi standar kualitas yang diharapkan.

Penelitian ini bertujuan untuk menganalisis kualitas butir soal tes *tashrif lughawi* menggunakan Rasch Model. Dengan melakukan analisis ini, diharapkan dapat ditemukan butir soal yang valid dan reliabel, serta mengidentifikasi butir soal yang memerlukan perbaikan. Penelitian ini juga bertujuan untuk memberikan rekomendasi bagi pengembang tes dalam merancang instrumen evaluasi yang lebih baik di masa depan.

Manfaat dari penelitian ini diharapkan tidak hanya dirasakan oleh praktisi pendidikan dan pengembang tes, tetapi juga oleh para peneliti dalam bidang evaluasi pendidikan bahasa. Hasil penelitian ini dapat digunakan untuk meningkatkan kualitas instrumen tes *tashrif lughawi*, sehingga evaluasi terhadap kemampuan peserta didik dapat dilakukan dengan lebih akurat dan konsisten. Selain itu, penelitian ini juga diharapkan dapat memberikan kontribusi dalam pengembangan teori dan praktik evaluasi pendidikan bahasa Arab secara umum.

Dalam upaya mencapai tujuan tersebut, penelitian ini akan menggunakan pendekatan Rasch Rasch untuk menganalisis data dari hasil tes *tashrif lughawi* yang telah dikumpulkan. Hasil analisis ini akan memberikan gambaran tentang kualitas butir soal dan kesesuaiannya dengan kemampuan peserta didik. Berdasarkan temuan penelitian, akan diberikan rekomendasi untuk perbaikan butir soal dan pengembangan instrumen tes yang lebih baik di masa mendatang.

2. Literature Review

Analisis suatu instrumen tes penting dilakukan karena mampu meningkatkan kemampuan pemecahan masalah melalui soal-soal tes yang mengukur berpikir tingkat tinggi yang bermuara pada pencapaian hasil belajar yang maksimal bagi siswa (Ismaulina, Yola Safitri, Aulia Widya Agustina, 2023). Pada konteks pendidikan, Salah satu cara dalam menganalisis suatu instrumen tes adalah menggunakan model Rasch, beberapa penelitian terdahulu telah membuktikan bahwa analisis instrumen soal menggunakan model Rasch dapat menghasilkan data yang konkrit

dalam mengevaluasi kualitas butir soal (Hayati & Lailatussaadah, 2016; Maulana & Aroyandini, 2024). Pada analisis soal asesmen bidang Ilmu Pengetahuan Sosial, instrumen memiliki kesesuaian dengan karakteristik siswa dan soal memiliki tingkat kesulitan yang bervariasi (Muharromah & Humaisi, 2020). Pada kualitas butir soal Matematika, ditemukan bahwa sebagian besar butir soal memenuhi kriteria kualitas yang baik dan menunjukkan bahwa soal memiliki tingkat kesulitan yang bervariasi dan dapat digunakan untuk mengukur kemampuan siswa (Hanna & Retnawati, 2022).

Materi tashrif lughawi merupakan materi yang krusial dalam pembelajaran Bahasa Arab, karena siswa harus benar-benar memahami setiap perubahan yang ada pada setiap dhamir (Huda, 2016). Pembelajaran Bahasa Arab pada materi tashrif lughawi harus memiliki instrument tes yang tepat agar dapat mengukur sejauh mana siswa dapat memahami materi (Muklason et al., 2023). Namun, penelitian yang menggunakan Model Rasch untuk menganalisis butir soal tes kognitif Bahasa Arab masih relatif terbatas. Penelitian yang dilakukan dalam pengembangan instrumen tes kemampuan Bahasa Arab bagi mahasiswa iain ponorogo menggunakan Metode penelitian R&D dengan model ADDIE (Ba'ul Chusna & Niswatul Fauziah, 2021) dalam penelitian. Analisis soal tes kompetensi kebahasaan pada Bahasa Arab di Madrasah Tsanawiyah memakai pendekatan kualitatif deskriptif analisis (Fauzi, 2023). Pada analisis daya beda tes hasil belajar Bahasa Arab siswa sma mamba'ul hikmah paron ngawi menggunakan metode deskriptif kuantitatif dalam analisis instrumen tes (Muslimah & Widiyanti, 2023).

Beberapa penelitian terdahulu tidak menggunakan model Rasch dalam menganalisis instrument tes untuk mengukur kognitif siswa dalam pembelajaran bahasa arab khususnya pada materi tashrif lughawi. Oleh karenanya, penelitian ini dilakukan untuk memberikan warna baru dalam praktik analisis instrument tes pada pembelajaran bahasa arab khususnya pada materi tashrif lughawi.

3. Method

Satu set soal sebagai instrumen dibangun untuk mengukur aspek

kognitif/kemampuan *Tashrif Lughawi* peserta didik pada mata pelajaran Bahasa Arab. Penelitian ini menganalisis kualitas butir soal dari aspek validitas, realibilitas dan tingkat kesukaran soal dari sejumlah 20 soal. Instrumen dalam bentuk soal objektif digunakan untuk penelitian ini dengan penskoran 1 bagi jawaban benar dan 0 bagi jawaban salah. Penelitian ini melibatkan peserta didik jenjang Madrasah Tsanawiyah sebagai responden (n=51). Data penelitian ini dianalisis menggunakan *Rasch Model* berbantuan aplikasi Winsteps versi. Data penelitian dilihat dengan teliti terlebih dahulu sebelum dientry dalam Winsteps dan selanjutnya dianalisis secara *Rasch Model* untuk mendapatkan hasil validitas, reliabilitas dan kesukaran soal.

4. Result and Discussion

Hasil penelitian ini menganalisis dan membahas dari beberapa aspek penting. Untuk memastikan suatu instrumen berkualitas, maka instrumen tersebut harus mampu memenuhi beberapa kriteria pengukuran, diantaranya validitas, reliabilitas dan tingkat kesukaran soal.

a. Validitas Item (Soal)

Validitas suatu instrument dapat diidentifikasi melalui analisis output dari aplikasi Winsteps. Output utama yang harus diidentifikasi adalah output *item polarity* untuk melihat koefisiensi korelasi yang dikenal dengan *Point Measure Correlation Coefficient* (PTMEA CORR). Nilai lainnya yang turut dilihat adalah pemetaan item dan responden dan indeks reliabilitas individu dan item.

(1) Item Polarity

Analisis polarity dilakukan untuk menunjukkan item-item yang dianalisis bergerak dalam satu arah yang dimaksud oleh konstruk yang diukur. Hasil indeks skor positif yang dihasilkan dari analisis *polarity* menunjukkan bahwa semua item soal berfungsi ke arah yang selaras untuk konstruk yang dibangun. *Point Measure Correlation Coefficient* (PTMEA CORR) digunakan untuk menentukan *item polarity*.

Jika skor PTMEA CORR adalah tinggi, maka item lebih berupaya untuk membedakan kemampuan antara responden. Sementara skor negatif (-) atau nol (0) menunjukkan item atau responden yang diperoleh bertentangan dengan variabel atau konstruk. Jika skor PTMEA CORR kurang dari 0.30 maka item dinyatakan longgar (Linacre, 2010). Hasil analisis *item polarity* dapat dilihat pada **Gambar 1** berikut.

TABLE 26.1 Data Soal Tashrif lughawi Artikel2 2. ZOU109WS.TXT Jul 2 2024 14:55
 INPUT: 51 PERSON 20 ITEM REPORTED: 51 PERSON 20 ITEM 2 CATS WINSTEPS 5.3.0.0
 PERSON: REAL SEP.: 1.97 REL.: .80 ... ITEM: REAL SEP.: 3.14 REL.: .91
 ITEM STATISTICS: CORRELATION ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	ITEM
10	28	51	-.81	.33	1.60	3.79	1.83	2.68	.12	.50	47.1	72.0	SOAL_10
4	29	50	-.95	.33	1.53	3.42	1.64	2.02	.16	.49	48.0	72.0	SOAL_4
2	41	51	-2.34	.38	1.34	1.62	1.17	.46	.16	.36	78.4	80.8	SOAL_2
15	16	51	.56	.36	1.47	2.21	1.55	1.67	.26	.54	66.7	78.8	SOAL_15
14	9	51	1.61	.43	1.48	1.78	1.09	.34	.27	.48	78.4	84.7	SOAL_14
9	33	51	-1.35	.33	1.18	1.30	1.12	.46	.35	.46	68.6	72.7	SOAL_9
5	39	51	-2.06	.36	1.02	.19	.88	-.08	.38	.39	82.4	77.9	SOAL_5
11	13	51	.96	.38	1.20	.96	1.39	1.06	.40	.52	74.5	81.2	SOAL_11
19	9	51	1.61	.43	.93	-.20	.96	.11	.50	.48	86.3	84.7	SOAL_19
16	14	51	.82	.37	.95	-.17	.90	-.18	.56	.53	78.4	80.5	SOAL_16
1	36	51	-1.69	.34	.78	-1.58	.64	-.92	.57	.43	88.2	74.5	SOAL_1
12	14	51	.82	.37	.87	-.62	.75	-.67	.61	.53	86.3	80.5	SOAL_12
8	26	51	-.59	.33	.82	-1.27	.77	-.98	.62	.51	82.4	72.5	SOAL_8
13	14	51	.82	.37	.84	-.76	.67	-.98	.63	.53	86.3	80.5	SOAL_13
7	23	51	-.27	.33	.83	-1.13	.74	-1.18	.64	.53	82.4	73.5	SOAL_7
18	13	51	.96	.38	.83	-.79	.55	-1.34	.65	.52	78.4	81.2	SOAL_18
20	7	51	2.00	.46	.59	-1.66	.27	-1.43	.69	.45	90.2	87.6	SOAL_20
3	32	51	-1.24	.33	.61	-3.40	.49	-1.89	.71	.47	94.1	72.2	SOAL_3
17	9	51	1.61	.43	.58	-1.94	.30	-1.69	.73	.48	86.3	84.7	SOAL_17
6	25	51	-.49	.33	.54	-3.78	.45	-2.89	.80	.52	94.1	72.8	SOAL_6
MEAN	21.5	51.0	.00	.37	1.00	-.10	.91	-.27			78.9	78.3	
P. SD	10.7	.2	1.30	.04	.33	1.98	.43	1.37			12.6	5.0	

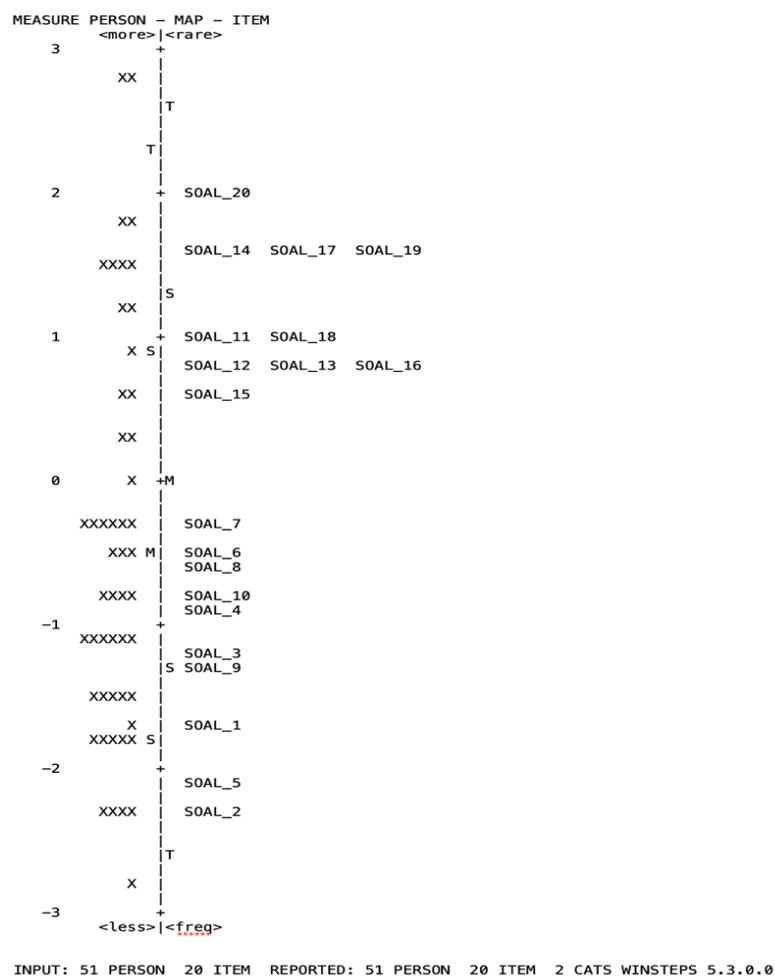
Gambar 1: PTMEA CORR Soal Tes Tashrif Lughawi

Berdasarkan **Gambar 1** bahwa semua item mempunyai skor PTMEA CORR adalah positif. Maka hal ini menunjukkan tidak terdapat pertentangan antara item dengan konstruk yang diukur. Selanjutnya merujuk kepada hasil OUTFIT yaitu skor *Mean Square* sebagaimana ditunjukkan **Gambar 1** bahwa hampir semua item memiliki skor *Mean Square* kurang dari 1.5 yang dapat diidentifikasi bahwa skor pengukuran dapat dinyatakan produktif kecuali terdapat 2 item yang skor *Mean Square* melebihi 1.5, yaitu item soal_10 (1.83), dan soal_4 (1.64). Namun demikian, skor ZSTD bagi kedua item tersebut adalah 2.6 untuk soal_10 dan 2.02 untuk soal_4. Maka keputusan bagi kedua item soal ini akan dilakukan kajian ulang untuk mengetahui permasalahan yang

ada sebelum item soal ini digugurkan dari instrumen tes pengetahuan *tashrif lughawi*.

(2) Pemetaan Item dan Responden

Pemetaan item dan responden dalam penelitian ini dilakukan untuk melihat kesesuaian setiap item soal yang dibangun dengan tahap kemampuan responden. Hasil pemetaan item dan responden sebagaimana ditunjukkan dalam **Gambar 2** berikut.



Gambar 2: Peta Sebaran Individu Item

Gambar 2 di atas menunjukkan peta sebaran individu item yang memplotkan kemampuan dan tingkat kesukaran item dalam logit yang sama. Berdasarkan **Gambar 2** bahwa skor rata-rata individu adalah -.46 yaitu kurang dari skor rata-

rata item yang nilainya adalah 0.0. Rata-rata kemampuan individu adalah lebih rendah daripada tingkat kesukaran item. Oleh karena itu, beberapa item soal perlu dikaji dan diperbaiki lagi untuk menurunkan tingkat kesukaran item agar setara antara kemampuan individu dan kesukaran item. Gambar ini menunjukkan 4 responden memiliki kemampuan sangat tinggi, yaitu responden ke-14, ke-20, ke-32 dan ke-34 memiliki kemampuan yang sangat tinggi (2.8 logits) dan terletak pada kedudukan paling teratas dalam map. Sedangkan responden ke-27 memiliki kemampuan yang sangat rendah (-2.8 logits) dan terletak pada kedudukan paling bawah dalam output map individu-item. Responden ke-14, ke-20, ke-32 dan ke-34 mampu menjawab hampir semua soal tes dengan benar. Hal ini dimungkinkan soal tes yang sangat mudah bagi mereka. Sedangkan responden ke-27 hanya mampu menjawab 2 soal dengan benar. Jika dilihat lebih rinci, kebanyakan responden memiliki kemampuan yang rendah dalam menjawab soal tes dengan benar, dimana jumlah jawaban yang benar dari jawaban mereka berada antara 2-9 soal dari 20 soal yang dites.

b. Reliabilitas Konstruk

Kualitas suatu alat ukur dalam aspek reliabilitas yang dilakukan pengukurannya melalui *Rasch Model* dapat ditunjukkan dengan menggunakan dua statistik, yaitu indeks reliabilitas sampel dan indeks pemisahan item (Effendi et. Al 2014). Reliabilitas bermaksud pada konsistensi jawaban responden terhadap instrument apabila diberikan atau diuji kepada responden yang berbeda namun memiliki ciri yang hampir sama (Adzhar et al., 2017). Hasil indeks reliabilitas sampel dan indeks pemisahan item adalah dipaparkan di dalam **Gambar 3** berikut.

TABLE 3.1 Data Soal Tashrif Lughawi Artikel2 2.. ZOU843WS.TXT Jul 3 2024 13:12
 INPUT: 51 PERSON 20 ITEM REPORTED: 51 PERSON 20 ITEM 2 CATS WINSTEPS 5.3.0.0

SUMMARY OF 51 MEASURED PERSON

	TOTAL	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
	SCORE				MNSQ	ZSTD	MNSQ	ZSTD
MEAN	8.4	20.0	-.46	.59	1.00	.02	.91	-.03
SEM	.6	.0	.19	.01	.04	.14	.06	.13
P.SD	4.4	.1	1.37	.07	.28	1.01	.45	.92
S.SD	4.4	.1	1.38	.07	.29	1.02	.45	.92
MAX.	18.0	20.0	2.77	.79	1.81	2.60	2.29	2.99
MIN.	2.0	19.0	-2.81	.53	.51	-2.25	.39	-2.07
REAL RMSE	.62	TRUE SD	1.22	SEPARATION	1.97	PERSON RELIABILITY	.80	
MODEL RMSE	.59	TRUE SD	1.24	SEPARATION	2.08	PERSON RELIABILITY	.81	
S.E. OF PERSON MEAN = .19								

PERSON RAW SCORE-TO-MEASURE CORRELATION = 1.00 (approximate due to missing data)
 CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .83 SEM = 1.79 (approx)
 STANDARDIZED (50 ITEM) RELIABILITY = .92

SUMMARY OF 20 MEASURED ITEM

	TOTAL	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
	SCORE				MNSQ	ZSTD	MNSQ	ZSTD
MEAN	21.5	51.0	.00	.37	1.00	-.10	.91	-.27
SEM	2.5	.1	.30	.01	.08	.45	.10	.31
P.SD	10.7	.2	1.30	.04	.33	1.98	.43	1.37
S.SD	11.0	.2	1.34	.04	.34	2.03	.44	1.41
MAX.	41.0	51.0	2.00	.46	1.60	3.79	1.83	2.68
MIN.	7.0	50.0	-2.34	.33	.54	-3.78	.27	-2.89
REAL RMSE	.40	TRUE SD	1.24	SEPARATION	3.14	ITEM RELIABILITY	.91	
MODEL RMSE	.37	TRUE SD	1.25	SEPARATION	3.37	ITEM RELIABILITY	.92	
S.E. OF ITEM MEAN = .30								

ITEM RAW SCORE-TO-MEASURE CORRELATION = -1.00 (approximate due to missing data)
 Global statistics: please see Table 44.
 UMEAN=.0000 USCALE=1.0000

Gambar 3: Reliabilitas Individu dan Item

Gambar 3 di atas menunjukkan indeks reliabilitas item yang dibangun adalah tinggi, yaitu 0.91. Sementara indeks reliabilitas sampel adalah 0.80 yang merupakan bahwa nilai ini berada pada tahap yang sangat baik dan dapat diterima (Sekaran & Bougie, 2016; Bond & Fox, 2015). Adapun indeks pemisahan item yang diperoleh adalah 3.14. Nilai ini menunjukkan bahwa item-item soal yang dibangun untuk mengukur kemampuan *Tashrif Lughawi* peserta didik dapat dibagi kepada 4 tahap kemampuan responden yang terlibat. Nilai ini dapat diterima disebabkan indeks pemisahan melebihi dari 2 (Fox & Jones, 1998). Indeks pemisahan sampel adalah 1.97 yang menunjukkan terdapat dua jenis tahap responden yang terlibat. Nilai ini masih lemah menurut (Fisher, 2007). Kandungan pengukuran reliabilitas

soal tes pengetahuan Tashrif Lughawi dapat dilihat dalam **Tabel 1** berikut.

Tabel 1: Isi Pengukuran Reliabilitas Soal Tes Pengetahuan Tashrif Lughawi

Konstruk			Jumlah Item (Soal)	Nilai Reliabilitas, α	
				Item	Responden
Instrumen	Tes	<i>Tashrif Lughawi</i>	20	0.91	0.80

5. Conclusion

Penelitian ini menguji kualitas butir soal tes pengetahuan *Tashrif Lughawi* menggunakan analisis *Rasch Model*. Butir soal *Tashrif Lughawi* dibangun merujuk kurikulum dan dilakukan pengujian kepada peserta didik jenjang Madrasah Tsanawiyah. Hasil analisis *Rasch Model* dari segi item polarity dapat dikatakan butir soal tes memiliki nilai yang valid dan instrument soal tes yang dibangun dapat mengukur pengetahuan *Tashrif Lughawi* peserta didik. Hasil uji reliabilitas soal tes adalah tinggi dan dapat diterima untuk mengukur pengetahuan *Tashrif Lughawi* peserta didik. Namun beberapa item soal perlu dikaji kembali untuk diperbaiki menjadi lebih baik. Peta sebaran individu dan item soal masih kurang mengukur kemampuan responden dalam menjawab soal tes. Sebaran soal tes masih banyak pada tingkat kesukaran yang tinggi. Dengan kata lain soal tes yang dibangun lebih sukar dan tidak dapat mengukur sejumlah responden yang tingkat kemampuan mereka yang rendah. Berkaitan dengan hal ini, maka penelitian ke depan diperlukan perbaikan pada item-item soal tes *Tashrif Lughawi* agar hasilnya akan lebih tepat dan konsisten.

Reference

- Adzhar, H., Aabdul Karim, A., & Sahrin, M. U. (2017). Pembangunan Instrumen Penerimaan E-Pembelajaran Pelajar Pascasiswazah Menggunakan Analisis Rasch. *Jurnal Pendidikan Malaysia*, 42(2), 147-155. <https://doi.org/10.17576/jpen-2017-42.02-08>
- Ba'ul Chusna, A., & Niswatul Fauziah, H. (2021). Pengembangan Instrumen Tes Kemampuan Bahasa Arab Bagi Mahasiswa Iain Ponorogo. *Ibriez : Jurnal Kependidikan Dasar Islam Berbasis Sains*, 6. <https://doi.org/10.21154/ibriez.v6i2.178>
- Bond, T., & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the*

- Human Science*. Routledge.
- Fattah, A. (2019). خصائص اللغة العربية ومزاياها. In *Alqalam* (Vol. 28, Issue 2, p. 261).
- Fauzi, S. (2023). Analisis Soal Tes Kompetensi Kebahasaan Pada PAS Bahasa Arab di Madrasah Tsanawiyah. *Ukazh: Journal of Arabic Studies*, 4(2), 635–848.
<https://doi.org/10.37274/ukazh.v4i2.873>
- Fisher, W. P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions*, 21(1095).
- Fox, C. M., & Jones, J. A. (1998). Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology*, 45(1), 30–45.
<https://doi.org/10.1037/0022-0167.45.1.30>
- Hanna, W. F., & Retnawati, H. (2022). Analisis Kualitas Butir Soal Matematika Menggunakan Model Rasch Dengan Bantuan Software Quest. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 11(4), 3695.
<https://doi.org/10.24127/ajpm.v11i4.5908>
- Hayati, S., & Lailatussaadah, L. (2016). Validitas Dan Reliabilitas Instrumen Pengetahuan Pembelajaran Aktif, Kreatif Dan Menyenangkan (Pakem) Menggunakan Model Rasch. *Jurnal Ilmiah Didaktika*, 16(2), 169.
<https://doi.org/10.22373/jid.v16i2.593>
- Huda, I. S. (2016). Belajar Bahasa Arab Dengan Metode Al-Kasysyaf : Memberantas Buta Makna Dalam Membaca. *Prosiding Konferensi Nasional Bahasa Arab II*, 146–154.
- Ismaulina, Yola Safitri, Aulia Widya Agustina, R. D. (2023). *Pengembangan Instrumen Tes Hasil Belajar Kelompok 3*.
- Linacre, J. M. (2010). *A user's guide to WINSTEP: Rasch Model computer programs, Program manual 3.680*. WINSTEPS.com.
- Matondang, Z. (2009). Validitas dan Reliabilitas Suatu Instrumen Penelitian. *JURNAL TABULARASA PPS UNIMED*, 6(1), 1510–1515.
<https://doi.org/10.4028/www.scientific.net/AMM.496-500.1510>
- Maulana, S., & Aroyandini, E. N. (2024). Analisis Model Rasch dalam Pengembangan Kuesioner Validasi Teoritik Butir Soal Pilihan Ganda Instrumen Tes Diagnostik Abstrak A . PENDAHULUAN Kesalahpahaman masih menjadi momok masalah dalam dunia pendidikan . Alasannya adalah bahwa pemahaman konsep yang. *Jurnal Konatif: Jurnal Ilmiah Pendidikan*, 1(2).
- Muharromah, M. M., & Humaisi, S. (2020). Analisis Butir Soal Penilaian Ujian Semester Gasal Mata Pelajaran IPS di MTs Darul Muna Ponorogo. *ASANKA: Journal of Social Science And Education*, 1(2), 102–114. <https://doi.org/10.21154/asanka.v1i2.2250>
- Muklason, A., Riksakomara, E., Mahananto, F., Djunaidy, A., Vinarti, R. A., Anggraeni, W., Nurita, R. T., Utamima, A., Basara, N. R., Nuriman, M. A., Jayanty, A., Khofifah, N., Laila, N., Mahatta, H. M., Ananda, M. F., Wibowo, T. R., Hisyam, M., & Kamil, M. Z. A. (2023). Pembuatan Media Pembelajaran Digital Interaktif untuk Materi Pembelajaran Bahasa Arab (Nahwu dan Shorof) untuk Santri Milenial. *Sewagati*, 7(3). <https://doi.org/10.12962/j26139960.v7i3.505>
- Muslimah, M., & Widiyanti, A. (2023). Analisis Daya Beda Tes Hasil Belajar Bahasa Arab Siswa SMA Mamba'ul Hikmah Paron Ngawi. *Al-Mu'Arrib: Journal of Arabic Education*, 3(2), 13–23. <https://doi.org/10.32923/al-muarrib.v3i2.3594>

- Putriyani, S. N. U., & Qalbah, I. N. (2023). Menyingkap Polemik Historisitas Tafsir Corak Lughawi. *Jurnal Iman Dan Spiritualitas*, 3(1), 43–50.
<https://doi.org/10.15575/jis.v3i1.21802>
- Samah, R., Isahak, A. B., Mohd Nathir, K. A. Bin, Abdul Majid, M. Y. Bin, Muhammad Nawawi, M. A. A. Bin, & Ahmad, H. Bin. (2023). Kepentingan Bahan Sastra dalam Pembelajaran Bahasa Arab Sebagai Bahasa Asing. *Malaysian Journal of Social Sciences and Humanities (MJSSH)*, 8(10), e002559.
<https://doi.org/10.47405/mjssh.v8i10.2559>
- Sekaran, U., & Bougie, R. (2016). *Research Methods for Business: A Skill Building Approach*. John Wiley & Sons.
- Sumintono, B., Islam, U., Indonesia, I., Widhiarso, W., & Mada, U. G. (2014). *Aplikasi Model RASCH Untuk Penelitian Ilmu-Ilmu Sosial*. November.
- Supandi, S., & Farikhah, L. (2016). Analisis Butir Soal Matematika Pada Instrumen Uji Coba Materi Segitiga. *JIPMat*, 1(1), 71–78.
<https://doi.org/10.26877/jipmat.v1i1.1085>