

Analisis Validitas Dan Reliabilitas Instrumen Penilaian Kemampuan Pemahaman Guru SD Pada Aspek Pembelajaran *Teaching At The Right Level* (TARL)

**Jurnal Insan
Peduli
Pendidikan
(JIPENDIK)**

Halaman 17-23

Firdaus¹, Arif Hidayad²

Research paper
Education

¹Pasca Sarjana, Pendidikan MIPA, Universitas Indraprasta PGRI, Jakarta, Indonesia

²Pendidikan Matematika, STKIP Taman Siswa Bima, Bima, Indonesia

Abstract

This study aims to analyze the validity and reliability of an assessment instrument designed to evaluate elementary school teachers' understanding of the Teaching at the Right Level (TaRL) approach. TaRL is a learning strategy that adapts instruction based on students' actual ability levels, rather than their grade or age. In its implementation, teachers' understanding of this approach is crucial; therefore, a valid and reliable assessment instrument is needed to measure it. This research uses a quantitative approach with a survey method and analyzes data from 214 elementary school teachers in Bima Regency, West Nusa Tenggara (NTB). The instrument consists of 13 multiple-choice items analyzed using the Rasch Model through the Jamovi application. The results show that the instrument has a reliability score of 0.440; however, all items meet the validity criteria based on infit and outfit values. Further analysis indicates that the distribution of item difficulty levels falls within the moderate category, with the exception of one item being the most difficult and one item being the easiest. The Wright Map and Item Characteristic Curve (ICC) confirm that most of the items possess appropriate measurement characteristics. This study contributes to the evaluation of level-based differentiated learning instruments and supports the development of teachers' capacity in implementing the TaRL approach.

Article Info

Article History:

Received 06/06/2025

Revised 24/06/2025

Accepted 30/06/2025

Available online

30/06/2025



Keywords:

Validitas, Reliabilitas, Instrumen Penilaian, Guru SD, *Teaching at the Right Level* (TaRL)

Corresponding Author:

Firdaus

Email: firdaus24031999@gmail.com

© The Author(s) 2025

JIPENDIK, Vol 3, No. 1, 2025
pp. 17-23

ISSN 3031-7231 (media
online)



CC BY: This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use.

Abstrak

Penelitian ini bertujuan untuk menganalisis validitas dan reliabilitas instrumen penilaian pemahaman guru sekolah dasar terhadap pendekatan *Teaching at the Right Level* (TaRL). Pendekatan TaRL merupakan strategi pembelajaran yang menyesuaikan pengajaran berdasarkan level kemampuan aktual siswa, bukan berdasarkan kelas atau usia. Dalam implementasinya, pemahaman guru terhadap pendekatan ini menjadi krusial, sehingga diperlukan instrumen penilaian yang valid dan reliabel untuk mengukurnya. Penelitian ini menggunakan pendekatan kuantitatif dengan metode survei dan menganalisis data dari 214 guru SD di Kabupaten Bima, NTB. Instrumen terdiri dari 13 butir soal pilihan ganda yang dianalisis menggunakan Model Rasch melalui aplikasi Jamovi. Hasil penelitian menunjukkan bahwa instrumen memiliki reliabilitas rendah (0,440), namun seluruh butir soal memenuhi kriteria validitas berdasarkan nilai infit dan outfit. Analisis lebih lanjut menunjukkan distribusi tingkat kesukaran butir berada dalam kategori sedang, dengan pengecualian pada satu butir paling sulit dan satu butir paling mudah. Wright Map dan kurva ICC mengonfirmasi bahwa sebagian besar butir instrumen memiliki karakteristik pengukuran yang sesuai. Penelitian ini memberikan kontribusi dalam evaluasi instrumen pembelajaran diferensiatif berbasis level serta pengembangan kapasitas guru dalam implementasi pendekatan TaRL.

Kata Kunci: Validitas, Reliabilitas, Instrumen Penilaian, Guru SD, Model Rasch

Pendahuluan

Peningkatan mutu pendidikan dasar merupakan salah satu fokus utama dalam menciptakan generasi yang cerdas dan kompetitif. Salah satu pendekatan yang banyak mendapat perhatian dalam beberapa tahun terakhir adalah *Teaching at the Right Level* (TaRL), yaitu strategi pembelajaran yang berorientasi pada kemampuan aktual siswa, bukan semata berdasarkan usia atau jenjang kelas. Pendekatan ini mengubah paradigma pembelajaran konvensional dengan menyesuaikan pengajaran berdasarkan tingkat kemampuan aktual peserta didik (Indartiningih et al., 2023). Salah satu indikator keberhasilan implementasi TaRL sangat bergantung pada pemahaman guru terhadap konsep pendidikan berdiferensiasi sesuai kebutuhan siswa (Hidayad et al., 2025). Oleh karena itu, diperlukan instrumen penilaian yang valid dan reliabel untuk mengukur pemahaman guru secara akurat dan konsisten.

Studi sebelumnya menunjukkan bahwa pelatihan TaRL berdampak positif pada peningkatan pemahaman guru terhadap metode diferensiasi pembelajaran (Ulfah et al., 2023). Namun, sebagian besar penelitian masih terfokus pada dampak intervensi pelatihan terhadap perubahan praktik mengajar, belum banyak yang secara khusus mengkaji kualitas alat ukur pemahaman guru itu sendiri. Dengan kata lain, terdapat *gap riset* dalam pengembangan dan validasi instrumen untuk mengevaluasi pemahaman guru terhadap pendekatan TaRL, terutama pada tingkat sekolah dasar. Hal ini menjadi penting karena instrumen yang tidak valid atau tidak reliabel dapat menghasilkan data yang menyesatkan dan berdampak pada pengambilan kebijakan pendidikan yang tidak tepat (Sugiharni, 2017).

Validitas dan reliabilitas merupakan dua aspek penting dalam evaluasi instrumen pengukuran. Validitas menunjukkan sejauh mana suatu instrumen benar-benar mengukur apa yang seharusnya diukur, sedangkan reliabilitas merujuk pada konsistensi hasil pengukuran dari waktu ke waktu (S. Azwar, 2014; Al Hakim et al., 2021). Dalam konteks pemahaman guru terhadap TaRL, aspek konseptual, prosedural, dan aplikatif menjadi domain utama yang harus tercakup oleh instrumen. Oleh sebab itu, perlu dilakukan pengujian validitas isi dan konstruk serta reliabilitas instrumen dengan metode yang kuat secara metodologis.

Penelitian ini hadir untuk menjawab kebutuhan tersebut dengan menganalisis validitas dan reliabilitas instrumen penilaian pemahaman guru SD terhadap pendekatan TaRL, menggunakan

pendekatan *Item Response Theory* (IRT), khususnya Model Rasch. Model ini memungkinkan pengukuran yang lebih objektif, dengan memperhatikan hubungan antara karakteristik butir soal dan kemampuan responden secara simultan. Selain itu, Rasch Model memberikan estimasi parameter yang invariant, serta alat visualisasi seperti Wright Map dan kurva ICC yang dapat memperkuat analisis kualitas instrumen (Aziz, 2015; Hadiyanti et al., 2024)

Penelitian ini memiliki *originalitas* pada dua hal. Pertama, fokus utamanya bukan pada efek pelatihan atau penerapan TaRL, melainkan pada pengembangan dan validasi instrumen pengukuran yang ditujukan khusus untuk mengevaluasi pemahaman guru secara kuantitatif. Kedua, penggunaan Model Rasch dalam konteks ini masih relatif jarang dilakukan, padahal pendekatan ini memberikan keunggulan dalam pengujian validitas konstruk dan unidimensionalitas instrumen pendidikan. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi ilmiah dalam pengembangan instrumen pendidikan yang berkualitas dan mendukung pengambilan keputusan berbasis data. Selain itu, apabila instrumennya memiliki indikator yang jelas, maka akan mampu menggali data sesuai dengan tujuan ukurnya (Hidayad & Artikel, 2017).

Secara umum, tujuan dari penelitian ini adalah untuk menganalisis kualitas instrumen penilaian pemahaman guru terhadap pendekatan TaRL melalui uji validitas dan reliabilitas menggunakan Model Rasch. Secara spesifik, penelitian ini bertujuan untuk: (1) menilai sejauh mana validitas butir soal dalam instrumen berdasarkan nilai *infit* dan *outfit*, (2) mengevaluasi reliabilitas instrumen secara keseluruhan, (3) menganalisis tingkat kesukaran dan kecocokan butir, serta (4) mengeksplorasi distribusi kemampuan responden dan kecocokan antara butir dan responden menggunakan Wright Map dan kurva ICC.

Dengan hasil analisis ini, penelitian diharapkan mampu memberikan kontribusi pada pengembangan instrumen evaluasi pembelajaran yang mendukung pendekatan diferensiasi berbasis level. Selain itu, penelitian ini juga dapat menjadi rujukan bagi institusi pendidikan dan pengambil kebijakan dalam merancang pelatihan atau asesmen pemahaman guru yang berbasis pada pendekatan TaRL secara lebih terstandar dan terpercaya

Metodologi Penelitian

Metode penelitian yang digunakan adalah metode kuantitatif dengan pendekatan survey. Instrumen yang sudah dikembangkan berbentuk soal pilihan ganda dengan 4 pilihan jawaban sebanyak 13 butir soal. Soal yang dibentuk terdiri dari beberapa indikator yakni: (1) asesmen kemampuan awal siswa, (2) pengelompokan berdasarkan level belajar, (3) strategi pembelajaran berbasis level, (4) monitoring kemajuan siswa, (5) refleksi dan penyesuaian pembelajaran. Data dikumpulkan dari sejumlah 214 guru SD di kabupaten Bima, propinsi NTB yang mengikuti pelatihan dan instrumen dibuat dalam bentuk google form (online). Analisis butir instrumen menggunakan *Model Rasch* untuk melihat validitas dan reliabilitas instrumen menggunakan aplikasi analisis data kuantitatif JAMOVI.

Hasil dan Pembahasan

Berdasarkan hasil analisis butir dengan menggunakan Rasch Model, maka hal-hal yang dibahas pada hasil penelitian ini adalah: reliabilitas instrumen, tingkat kesukaran butir, fitnya butir, *wright map*, kurva ICC per butir, dan *item outfit plot*. Berikut pembahasannya secara rinci pada masing-masing item.

Reliabilitas Instrumen

Dari hasil analisis menunjukkan bahwa skor reliabilitas yang diperoleh adalah sebesar 0,440. Jika dikaitkan dengan kriteria reliabilitas yang ditetapkan maka reliabilitas instrumen berada pada kategori "Rendah" yang ditunjukkan pada Tabel 1. Skor reliabilitas ini dapat ditingkatkan, apabila mengeluarkan butir-butir yang memiliki korelasi rendah dalam analisis.

Tabel 1. Reliabilitas Instrumen

Model Fit		
	Person Reliability	
scale	0.440	

Tingkat kesukaran butir dan Item Fit

Pada Tabel 2 menunjukkan bagaimana tingkat kesukaran pada setiap butir soal. Pada analisis tes klasik, terlihat bahwa proporsi (proportion) menjawab benar pada setiap butir soal berada pada rentang skor 0,238 – 0,939. Semakin tinggi skornya maka semakin “mudah” butir tersebut dijawab oleh peserta tes, sebaliknya skor yang rendah menunjukkan bahwa butir semakin “sulit” untuk dijawab oleh peserta. Tabel 2 menginformasikan bahwa butir 3 (But_3) adalah butir soal yang paling “sulit” dari butir-butir lain, sedangkan butir soal yang paling mudah yakni butir 13 (But_13). Hal ini juga diperkuat dari hasil analisis Rasch Model (skor Measure) yang menyatakan bahwa butir 3 memiliki skor sebesar 0,1260 yang lebih tinggi dibanding butir-butir yang lain. Sedangkan pada butir 13 memiliki skor sebesar -2,90.

Selanjutnya, untuk mengetahui butir ini fit (valid), maka skor yang diperhatikan ada pada skor Outfit. Butir yang fit (valid) adalah butir yang terletak pada rentang skor Outfit 0,50 s.d 1,50. Pada tabel 2 terlihat bahwa semua butir berada pada rentang tersebut dan dapat dikatakan butir-butir soal berada pada kategori “fit (Valid)”

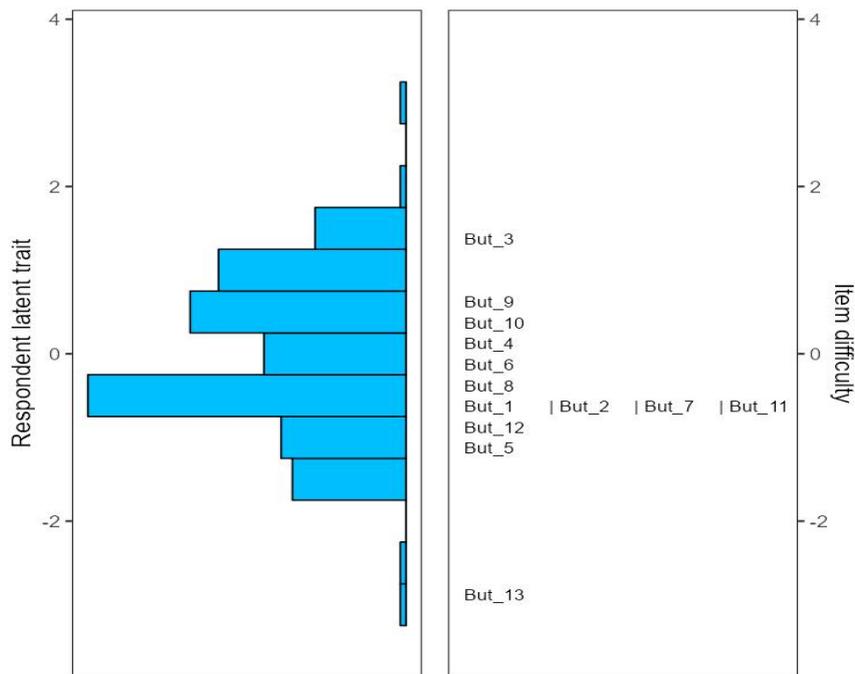
Tabel 2. Item Statistics setiap butir instrumen

Butir	Proportion	Measure	S.E.Measure	Infit	Outfit
But_1	0.650	-0.671	0.149	0.946	0.928
But_2	0.645	-0.649	0.149	0.966	0.965
But_3	0.238	1.260	0.166	1.078	1.161
But_4	0.472	0.124	0.143	1.017	1.026
But_5	0.743	-1.152	0.162	0.949	0.928
But_6	0.556	-0.245	0.144	1.041	1.047
But_7	0.650	-0.671	0.149	0.992	0.987
But_8	0.603	-0.454	0.146	0.953	0.942
But_9	0.388	0.499	0.146	1.082	1.099
But_10	0.407	0.414	0.145	1.048	1.056
But_11	0.617	-0.518	0.147	0.970	0.964
But_12	0.710	-0.974	0.157	0.962	0.951
But_13	0.939	-2.911	0.290	0.983	0.818

Note. Infit= Information-weighted mean square statistic; Outfit= Outlier-sensitive means square statistic.

Wright Map

Pada Wright Map menjelaskan tentang bagaimana sebaran kemampuan peserta tes dengan sebaran tingkat kesulitan butir soal seperti terlihat pada gambar 1.



Source: pribadi

Gambar 1. Wright Map pada butir instrumen

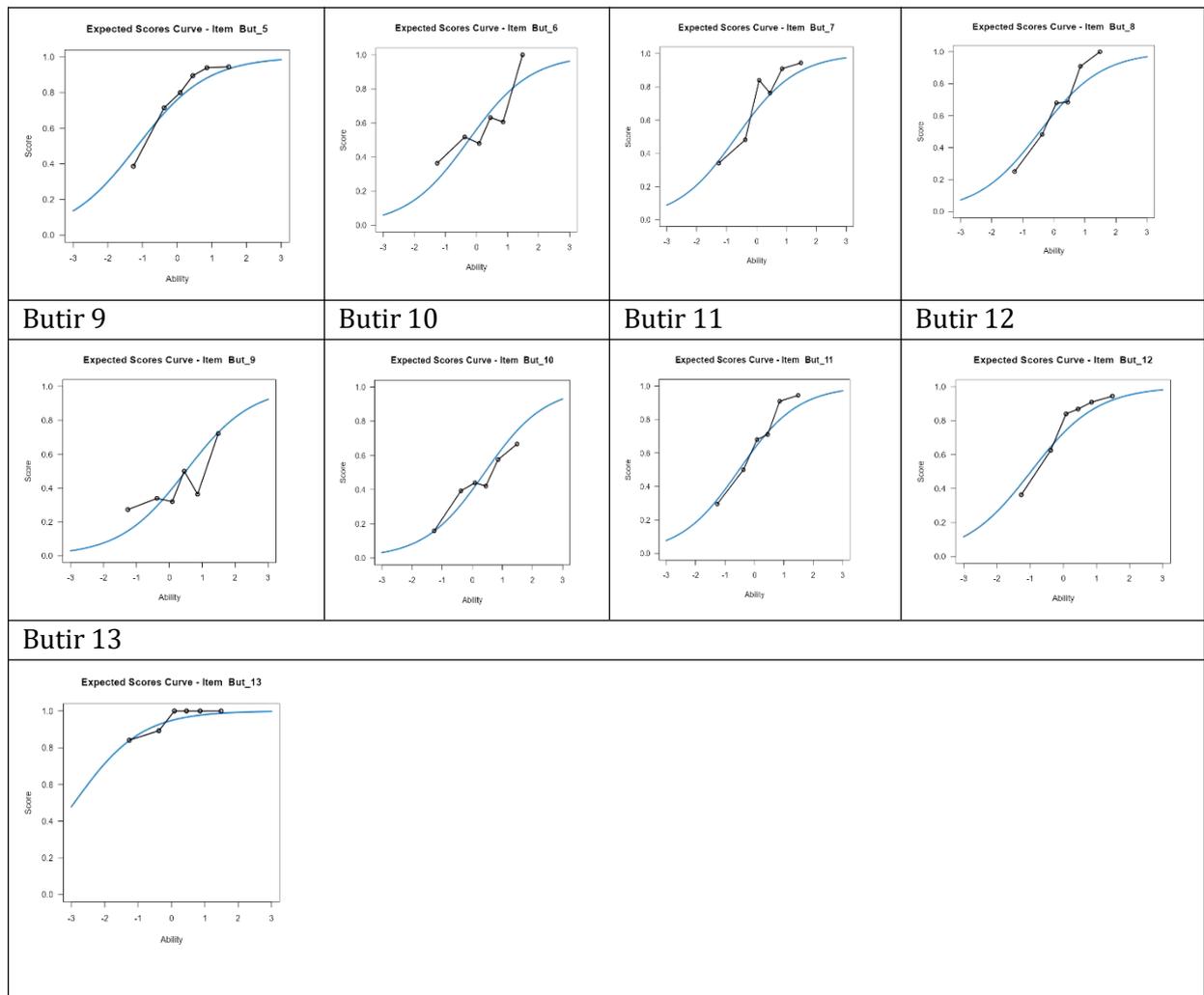
Pada gambar 1 dapat disampaikan bahwa kemampuan peserta tes mengumpul ditengah-tengah yakni berada pada kategori sedang. Begitupun dengan butir-butir instrumen terlihat mengumpul pada butir dengan kategori “sedang” untuk tingkat kesulitannya. Terdapat beberapa peserta tes yang memiliki kemampuan tinggi, namun tidak ada butir yang mampu mengukur kemampuan peserta tersebut

Kurva Item Characteristic Curve (ICC) tiap butir instrumen

Kurva ICC menunjukkan bagaimana karakteristik butir yang baik. Butir soal yang baik adalah yang membentuk pola seperti huruf S. Dari 13 butir terlihat bahwa hanya butir 13 yang tidak membentuk huruf S atau tidak mengarah pada kurva ICC yang diharapkan. Hal tersebut dapat dilihat pada Tabel 3.

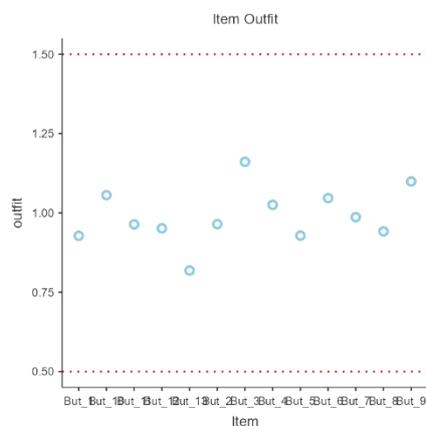
Tabel 3. Kurva ICC pada setiap butir instrumen

Butir 1	Butir 2	Butir 3	Butir 4
Butir 5	Butir 6	Butir 7	Butir 8



Item Outfit plot

Pada item outfit plot menunjukkan bahwa seluruh butir soal berada pada kategori rentang yang wajar atau nilai outfitnya tidak melebihi 1,25 dan tidak kurang dari 0,50 (Wahyuni, 2014), seperti terlihat pada gambar 2. Hal ini menunjukkan bahwa seluruh butir instrumen memenuhi valid pada Rasch Model.



Source: pribadi

Gambar 2. Item outfit Plot

Kesimpulan

Berdasarkan hasil analisis dan pembahasan, dapat disimpulkan bahwa instrumen penilaian pemahaman guru terhadap pendekatan TaRL yang diuji pada 214 responden menunjukkan reliabilitas yang tergolong rendah dengan skor 0,440. Nilai reliabilitas ini masih dapat ditingkatkan melalui penyempurnaan item, khususnya dengan mengevaluasi butir-butir yang memiliki korelasi rendah. Dari 13 butir soal, satu butir termasuk kategori sangat sulit, satu butir sangat mudah, dan 11 lainnya berada pada tingkat kesulitan sedang. Seluruh butir soal dinyatakan valid berdasarkan analisis fit (*infit* dan *outfit*) Model Rasch, dan mayoritas kurva ICC menunjukkan pola berbentuk huruf S, kecuali satu butir yang tidak sesuai. Wright Map juga menunjukkan kesesuaian antara tingkat kemampuan peserta dan kesulitan butir soal, meskipun ada celah dalam mengukur peserta dengan kemampuan sangat tinggi. Secara keseluruhan, instrumen telah menunjukkan validitas yang baik namun memerlukan perbaikan untuk meningkatkan reliabilitas agar lebih representatif dan akurat dalam mengukur pemahaman guru terhadap pendekatan TaRL.

Daftar Pustaka

- Al Hakim, R., Mustika, I., & Yuliani, W. (2021). Validitas Dan Reliabilitas Angket Motivasi Berprestasi. *FOKUS (Kajian Bimbingan & Konseling Dalam Pendidikan)*, 4(4), 263. <https://doi.org/10.22460/fokus.v4i4.7249>
- Aziz, R. (2015). Aplikasi Model RASCH dalam Pengujian Alat Ukur Kesehatan Mental di Tempat Kerja. *Psikoislamika: Jurnal Psikologi Dan Psikologi Islam*, 12(2), 29. <https://doi.org/10.18860/psi.v12i2.6402>
- Hadiyanti, H., Susongko, P., & Munadi. (2024). Pengembangan Instrumen Higher Order Thinking Skill Mata Pelajaran Matematika Dengan Rasch Model. *Journal Od Education Research*, 5, 402.
- Hidayad, A. (2017). Instrumen Asesmen Sikap Siswa Berbasis Konservasi pada Pembelajaran. 6(1), 30–38.
- Hidayad, A., Ghufroon, A., Hidayati, K., Larasati, C. N., Suratman, & Ashari, L. H. (2025). *What does the principal do to improve the reading skills of lower-grade students? Case Studies in Elementary Schools in Indonesia. Educational Process: International Journal*, 16. <https://doi.org/10.22521/edupij.2025.16.191>
- Indartiningsih, D., Mariana, N., & Subrata, H. (2023). Perspektif Global Dalam Implementasi Teaching At The Right Level(Tarl) Pada Pembelajaran Berdiferensiasi Pada Kurikulum Merdeka. *Jurnal Elementaria Edukasia*, 6(4), 1984–1994. <https://doi.org/10.31949/jee.v6i4.7547>
- S. Azwar. (2014). Penyusunan Skala Psikologi (Kedua). Pustaka Belajar.
- Sugiharni, G. A. D. (2017). Validitas Isi Instrumen Pengujian Modul Digital Matematika Diskrit Berbasis Open Source di STIKOM Bali. *E-Proceedings KNS&I STIKOM Bali*, 678–684.
- Ulfah, A., Fatmawati, L., Purnama, R. D., Pratama, F. Y., & Adityas, M. T. (2023). *TaRL-Based Differentiated Learning Model Training for Primary School Teachers in Independent Curriculum Implementation. Social, Humanities, and Educational Studies (SHES): Conference Series*, 6(3), 348–356. <https://doi.org/10.20961/shes.v6i3.82370>
- Wahyuni, K. M. dan S. (2014). Analisis Kemampuan Peserta Didik Dengan Model Rasch. *Seminar Nasional Evaluasi Pendidikan Tahun 2014*, 97–98.