



Analisis kualitas instrumen PAS pada mata pelajaran al-qur'an hadits menggunakan model rasch

Ulva Rahmi*

UIN Sjech M. Djamil
Djambek Bukittinggi,
INDONESIA

Razzaq Fikih Al Zikra

UIN Sjech M. Djamil
Djambek Bukittinggi,
INDONESIA

Rizky Illahi Yusnaidi

UIN Sjech M. Djamil
Djambek Bukittinggi,
INDONESIA

Miftahir Rizqa

UIN Sjech M. Djamil
Djambek Bukittinggi,
INDONESIA

Article Info

Article history:

Received: Jan 19, 2026

Revised: Feb 26, 2026

Accepted: March 12, 2026

Keywords:

Al-Qur'an Hadits, evaluasi pembelajaran; Exploratory Factor Analysis; Model Rasch; Reliabilitas instrumen, Validitas Konstruk.

Abstract

Evaluasi pembelajaran memiliki peran penting dalam memastikan ketercapaian tujuan pendidikan, namun kualitas evaluasi sangat bergantung pada validitas dan reliabilitas instrumen yang digunakan. Dalam praktiknya, masih banyak instrumen penilaian yang belum melalui analisis empiris secara komprehensif, sehingga berpotensi menghasilkan pengukuran yang bias. Penelitian ini bertujuan untuk menganalisis kualitas instrumen penilaian akhir semester pada mata pelajaran Al-Qur'an Hadits melalui pendekatan Model Rasch yang dipadukan dengan analisis faktor, yaitu Exploratory Factor Analysis (EFA) dan Confirmatory Factor Analysis (CFA). Penelitian menggunakan pendekatan kuantitatif deskriptif dengan melibatkan 253 peserta didik kelas VII di MTsN 2 Kota Bukittinggi. Instrumen yang dianalisis terdiri dari 50 butir soal pilihan majemuk, dengan data dikumpulkan melalui teknik dokumentasi. Analisis data dilakukan menggunakan perangkat lunak SPSS dan JASP untuk menguji validitas konstruk dan reliabilitas instrumen. Hasil penelitian menunjukkan bahwa instrumen memiliki karakteristik unidimensional berdasarkan analisis EFA, serta mengalami peningkatan kesesuaian model setelah dilakukan modifikasi pada tahap CFA, yang ditunjukkan oleh peningkatan nilai indeks fit dan penurunan nilai RMSEA. Selain itu, hasil uji reliabilitas menunjukkan nilai Cronbach's Alpha sebesar 0,791 yang mengindikasikan konsistensi internal yang baik. Dengan demikian, instrumen yang dianalisis memiliki kualitas yang memadai untuk digunakan dalam evaluasi pembelajaran. Penelitian ini menyimpulkan bahwa integrasi Model Rasch dengan EFA dan CFA memberikan pendekatan yang lebih komprehensif dalam mengevaluasi kualitas instrumen, serta disarankan agar pengembangan instrumen dilakukan secara sistematis dan mencakup berbagai ranah kemampuan peserta didik.

To cite this article: Rahmi, U., Al Zikra, R. F., Yusnaidi, R. I., & Rizqa, M. (2026). Analisis kualitas instrumen pas pada mata pelajaran al qur'an hadits menggunakan model rasch. *Al Qodiri: Jurnal Pendidikan, Sosial dan Keagamaan*, 24(1), 51-61.

INTRODUCTION

Evaluasi pembelajaran merupakan komponen fundamental dalam sistem pendidikan karena berperan sebagai dasar dalam menilai ketercapaian tujuan pembelajaran sekaligus menjadi landasan dalam pengambilan keputusan yang berbasis data. Dalam perspektif ini, evaluasi tidak hanya dipahami sebagai aktivitas administratif, tetapi sebagai proses ilmiah yang menuntut ketepatan pengukuran dan keakuratan interpretasi. Kualitas evaluasi pembelajaran sangat ditentukan oleh kualitas instrumen yang digunakan, karena instrumen tersebut menjadi alat utama dalam merepresentasikan kemampuan peserta didik secara empiris (Laila et al., 2024; Muryadi, 2017). Instrumen yang tidak memenuhi standar pengukuran yang baik berpotensi menghasilkan data yang bias, sehingga dapat mengarah pada interpretasi yang keliru terhadap capaian belajar peserta didik.

* Corresponding author:

Ulva Rahmi, UIN Sjech M. Djamil Djambek Bukittinggi, INDONESIA

ulvarahmi@uinbukittinggi.ac.id

Dalam praktik pendidikan, pengukuran hasil belajar, khususnya pada ranah kognitif, umumnya dilakukan melalui penilaian akhir semester dengan menggunakan instrumen pilihan majemuk. Instrumen ini banyak digunakan karena memiliki efisiensi dalam mengukur cakupan materi yang luas serta kemudahan dalam proses penskoran. Namun demikian, di balik keunggulan tersebut, instrumen pilihan majemuk menuntut kualitas konstruk yang tinggi agar mampu menghasilkan data yang valid dan reliabel. Instrumen yang baik harus mampu mengukur kompetensi yang relevan secara akurat dan konsisten, serta memiliki daya pembeda yang memadai untuk membedakan tingkat kemampuan peserta didik (Arikunto, 2021; Linn, 2006). Selain itu, kualitas instrumen juga perlu didukung oleh bukti validitas dan reliabilitas yang diperoleh melalui pengujian empiris yang sistematis, sehingga hasil pengukuran dapat dipercaya dan digunakan sebagai dasar pengambilan keputusan pendidikan yang tepat (Khumaedi, 2012; Sugiono et al., 2020). Tanpa adanya pengujian kualitas instrumen yang memadai, hasil evaluasi berpotensi tidak mencerminkan kondisi sebenarnya dari capaian belajar, sehingga dapat mengurangi validitas keputusan pendidikan yang diambil.

Meskipun pentingnya kualitas instrumen telah banyak disadari, permasalahan yang masih sering ditemukan di lapangan adalah penggunaan instrumen evaluasi yang belum melalui proses analisis empiris secara memadai. Dalam banyak kasus, instrumen disusun secara praktis tanpa melalui pengujian validitas, reliabilitas, maupun analisis karakteristik butir secara mendalam. Kondisi ini berimplikasi pada rendahnya kualitas instrumen yang digunakan dalam evaluasi pembelajaran. Instrumen yang tidak valid, tidak reliabel, atau memiliki daya pembeda rendah berpotensi menghasilkan bias dalam pengukuran dan tidak mampu merepresentasikan kemampuan peserta didik secara akurat (Al Hakim et al., 2021; Devi et al., 2022). Kondisi tersebut juga menunjukkan bahwa pengembangan instrumen sering kali belum didasarkan pada prosedur analisis psikometrik yang memadai, seperti pengujian konstruk melalui pendekatan statistik yang sistematis (Heri et al., 2017). Akibatnya, hasil evaluasi yang diperoleh tidak hanya kurang akurat, tetapi juga dapat mempengaruhi kualitas keputusan yang diambil oleh pendidik dalam merancang tindak lanjut pembelajaran.

Seiring dengan perkembangan ilmu pengukuran pendidikan, berbagai pendekatan analisis modern telah dikembangkan untuk meningkatkan kualitas evaluasi pembelajaran. Salah satu pendekatan yang banyak digunakan adalah Model Rasch sebagai bagian dari Item Response Theory (IRT). Model ini memungkinkan analisis yang lebih mendalam terhadap kualitas instrumen, khususnya dalam mengidentifikasi tingkat kesulitan butir, kemampuan responden, serta kesesuaian antara data empiris dengan model pengukuran (Azzahra et al., 2024; Scoulas et al., 2021; Yudha, 2023). Selain itu, pendekatan IRT juga memberikan keunggulan dalam menghasilkan pengukuran yang lebih objektif karena tidak bergantung pada sampel tertentu, sehingga mampu memberikan estimasi kemampuan yang lebih stabil dibandingkan teori tes klasik (Chan et al., 2021; Syafii et al., 2021).

Di sisi lain, analisis faktor juga menjadi pendekatan yang penting dalam menguji validitas konstruk suatu instrumen. Exploratory Factor Analysis (EFA) digunakan untuk mengidentifikasi struktur dimensi yang terbentuk dari data empiris, sedangkan Confirmatory Factor Analysis (CFA) digunakan untuk menguji kesesuaian model konstruk yang telah dirumuskan secara teoritis (Kline, 2023; Putri & Febrilia, 2024; Stapleton, 1997). Pendekatan ini juga banyak digunakan dalam berbagai penelitian untuk mengidentifikasi faktor-faktor laten yang mempengaruhi suatu fenomena secara empiris (Ariesa, 2019). Selain itu, pendekatan CFA banyak digunakan dalam penelitian pengukuran modern untuk menguji model konstruk yang kompleks serta memastikan bahwa indikator yang digunakan benar-benar merepresentasikan variabel laten yang diukur (Ghazali & Nordin, 2019; Prasetyo et al., 2020). Bahkan, pengujian model pengukuran juga dapat dilakukan melalui pendekatan congeneric, tau-equivalent, dan parallel model guna memastikan konsistensi konstruk secara lebih mendalam (Natanael & Novanto, 2020). Perkembangan analisis data dalam penelitian pendidikan menunjukkan bahwa pemanfaatan teknik statistik yang lebih kompleks dapat meningkatkan ketepatan interpretasi hasil evaluasi serta memperkuat validitas temuan penelitian (Santoso et al., 2022).

Meskipun berbagai pendekatan tersebut telah banyak digunakan dalam penelitian evaluasi pendidikan, sebagian besar kajian masih dilakukan secara parsial. Penelitian cenderung hanya menggunakan Model Rasch atau analisis faktor secara terpisah tanpa mengintegrasikan kedua

pendekatan tersebut dalam satu kerangka analisis yang komprehensif. Pendekatan yang terfragmentasi ini menyebabkan analisis kualitas instrumen belum sepenuhnya mampu menggambarkan hubungan antara karakteristik butir dengan struktur konstruk secara simultan. Selain itu, penelitian yang secara khusus mengkaji integrasi Model Rasch dengan analisis faktor dalam konteks pendidikan agama Islam, khususnya pada mata pelajaran Al-Qur'an Hadits di tingkat madrasah, masih relatif terbatas. Kondisi ini menunjukkan adanya kesenjangan dalam literatur yang memerlukan pendekatan analisis yang lebih komprehensif dan terintegrasi untuk menghasilkan evaluasi instrumen yang lebih akurat.

Model Rasch sebagai pendekatan dalam teori respons butir menawarkan kerangka pengukuran yang kuat untuk mengevaluasi kualitas instrumen secara lebih mendalam. Model ini memungkinkan peneliti untuk menganalisis tingkat kesulitan butir dan kemampuan responden dalam satu skala pengukuran yang sama, sehingga memberikan interpretasi yang lebih bermakna terhadap data yang dihasilkan (Boone, 2016; Yudha, 2023). Keunggulan utama Model Rasch terletak pada sifatnya yang independen terhadap sampel, sehingga hasil pengukuran yang diperoleh lebih objektif dan tidak dipengaruhi oleh karakteristik kelompok responden tertentu (Chan et al., 2021; Pratama, 2020). Selain itu, model ini juga mampu memberikan informasi penting terkait reliabilitas instrumen, pemisahan butir dan individu, serta unidimensionalitas konstruk yang diukur (Retnawati, 2016).

Berdasarkan kerangka konseptual dan kesenjangan penelitian tersebut, studi ini diarahkan untuk mengevaluasi kualitas instrumen penilaian akhir semester pada mata pelajaran Al-Qur'an Hadits melalui penerapan Model Rasch yang dipadukan dengan analisis faktor, yaitu Exploratory Factor Analysis (EFA) dan Confirmatory Factor Analysis (CFA). Integrasi pendekatan ini digunakan untuk menghasilkan evidensi empiris yang lebih komprehensif, baik pada level karakteristik butir maupun pada struktur konstruk yang mendasari instrumen, sehingga mampu memberikan gambaran kualitas instrumen secara lebih menyeluruh.

Secara teoretis, penelitian ini berkontribusi dalam pengembangan kajian evaluasi pembelajaran dengan menawarkan pendekatan analisis terpadu yang menghubungkan perspektif Item Response Theory dan analisis faktor dalam satu kerangka evaluasi instrumen yang saling melengkapi. Sementara itu, secara praktis, temuan penelitian ini diharapkan dapat menjadi rujukan bagi pendidik dalam menyusun instrumen evaluasi yang lebih akurat, valid, dan reliabel. Lebih lanjut, penelitian ini juga diharapkan dapat memperkuat praktik evaluasi pembelajaran berbasis data empiris, khususnya dalam konteks pendidikan agama Islam di lingkungan madrasah, sehingga mampu meningkatkan kualitas pengukuran capaian belajar peserta didik secara berkelanjutan.

METHOD

Penelitian ini menggunakan pendekatan kuantitatif dengan metode deskriptif kuantitatif yang bertujuan untuk menggambarkan secara sistematis karakteristik instrumen evaluasi pembelajaran berdasarkan data numerik yang diperoleh. Pendekatan ini dipilih karena memungkinkan analisis objektif terhadap kualitas instrumen melalui data empiris tanpa melakukan pengujian hubungan antar variabel. Sebagaimana dikemukakan oleh Unaradjan (2019), metode deskriptif kuantitatif berfokus pada pengumpulan dan analisis data dalam bentuk angka untuk memberikan gambaran menyeluruh mengenai fenomena yang diteliti. Dalam konteks penelitian ini, pendekatan tersebut digunakan untuk mengevaluasi kualitas instrumen penilaian akhir semester pada mata pelajaran Al-Qur'an Hadits, khususnya pada aspek validitas konstruk dan reliabilitas butir soal pilihan majemuk.

Penelitian dilaksanakan di MTsN 2 Kota Bukittinggi dengan melibatkan 253 peserta didik kelas VII yang telah mengikuti ujian akhir semester secara lengkap. Teknik pengambilan sampel menggunakan metode sensus atau total sampling, di mana seluruh populasi dijadikan sebagai sampel penelitian. Pemilihan teknik ini dilakukan untuk memastikan bahwa analisis instrumen didasarkan pada keseluruhan data respon peserta didik, sehingga hasil yang diperoleh lebih representatif terhadap kondisi sebenarnya. Data penelitian dikumpulkan melalui teknik dokumentasi, yaitu dengan menghimpun naskah soal, kunci jawaban, serta lembar jawaban peserta didik yang kemudian dikonversi ke dalam format data numerik untuk keperluan analisis.

Instrumen yang dianalisis dalam penelitian ini berupa 50 butir soal pilihan majemuk yang masing-masing terdiri dari empat opsi jawaban (A, B, C, dan D) dengan satu jawaban benar sebagai kunci. Instrumen ini digunakan untuk mengukur capaian hasil belajar peserta didik pada mata pelajaran Al-Qur'an Hadits. Analisis instrumen difokuskan pada dua aspek utama, yaitu validitas dan reliabilitas. Validitas konstruk dianalisis menggunakan pendekatan analisis faktor, yaitu Exploratory Factor Analysis (EFA) dan Confirmatory Factor Analysis (CFA), untuk mengidentifikasi struktur dimensi serta menguji kesesuaian model konstruk instrumen. Dalam analisis EFA, parameter yang diperhatikan meliputi nilai Kaiser-Meyer-Olkin (KMO) dan unidimensionalitas instrumen, sedangkan dalam analisis CFA digunakan indeks kesesuaian model seperti Root Mean Square Error of Approximation (RMSEA), nilai derajat kebebasan (df), serta indikator kesesuaian model lainnya. Sementara itu, reliabilitas instrumen dianalisis menggunakan koefisien Cronbach's Alpha, dengan kriteria bahwa instrumen dinyatakan reliabel apabila memiliki nilai koefisien $\geq 0,70$.

Selain analisis faktor, penelitian ini juga memanfaatkan pendekatan Model Rasch sebagai bagian dari teori respons butir untuk mengevaluasi karakteristik butir soal dan kemampuan responden secara simultan. Model Rasch digunakan untuk mengidentifikasi tingkat kesulitan butir, kesesuaian butir (item fit), serta distribusi kemampuan peserta didik dalam satu skala logit yang sama. Pendekatan ini memungkinkan evaluasi yang lebih objektif terhadap kualitas instrumen karena tidak bergantung pada karakteristik sampel tertentu. Seluruh proses analisis data dilakukan dengan bantuan perangkat lunak JASP dan SPSS, yang digunakan untuk mengolah data EFA, CFA, serta reliabilitas instrumen. Hasil analisis kemudian disajikan secara deskriptif untuk memberikan gambaran komprehensif mengenai kualitas instrumen penilaian akhir semester pada mata pelajaran Al-Qur'an Hadits sebagai alat evaluasi pembelajaran.

RESULTS AND DISCUSSION

Results

Hasil Uji Validitas Instrumen

Hasil Uji Hasil Uji Validitas Instrumen (EFA)

Analisis validitas instrumen dilakukan menggunakan pendekatan Exploratory Factor Analysis (EFA) untuk mengidentifikasi struktur faktor serta kecenderungan unidimensionalitas instrumen. Analisis ini bertujuan untuk mengelompokkan butir-butir soal berdasarkan faktor yang terbentuk dari data empiris.

Dalam penelitian ini, beberapa indikator utama yang digunakan dalam analisis EFA meliputi nilai Kaiser-Meyer-Olkin (KMO), Bartlett's Test, serta scree plot. Nilai KMO dan Bartlett's Test digunakan untuk menentukan kelayakan data dalam analisis faktor, sedangkan scree plot digunakan untuk mengidentifikasi jumlah faktor yang terbentuk berdasarkan distribusi eigenvalue.

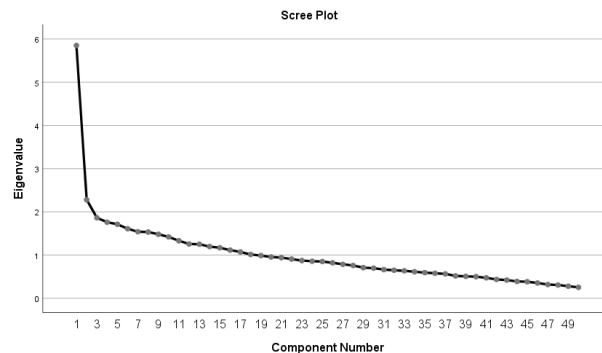
Pengolahan data dilakukan dengan menggunakan perangkat lunak SPSS dan JASP untuk memastikan keakuratan hasil analisis. Tahap awal dilakukan dengan menguji nilai KMO dan Bartlett's Test untuk mengetahui apakah data memenuhi syarat untuk dilanjutkan ke analisis faktor. Hasil pengujian tersebut disajikan pada tabel berikut.

Tabel 1. Tabel KMO dan Bartlett's Test Menggunakan SPSS 25

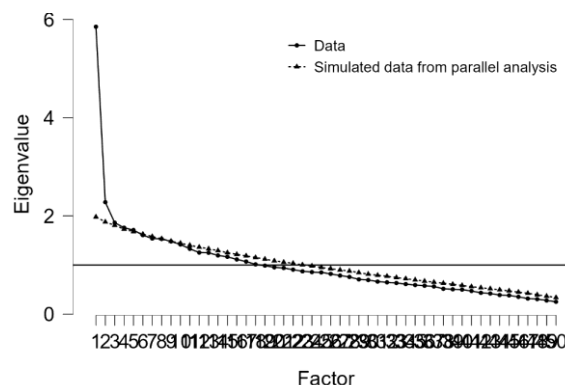
KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,694
Bartlett's Test of Sphericity	Approx. Chi-Square	2413,279
	Df	1225
	Sig.	,000

Dari data diatas dapat dilihat bahwa nilai KMO MSA Adalah 0,694 yang dibulatkan menjadi 0,70 dan nilai Barlett's Test Adalah 2413,279 dengan df atau derajat kebesaran sebesar 1225 dengan taraf signifikan 0,00 yang menandakan bahwa konstruk pada instrument bisa dilakukan uji analisis factor.

Setelah didapatkan hasil KMO dan Barlett's Test maka masuk ketahap uji EFA instrumen. Dalam pengujian ini, yang akan diamati adalah curaman dari grafik Eigenvalue. Grafik tersebut akan digunakan untuk menguji teori respon butir unidimensional dari instrumen. Menurut Hambleton dan Swaminathan Unidimensional berarti setiap butir test hanya mengukur satu kemampuan atau mengukur dimensi tunggal (Hambleton & Swaminathan, 2013). Hasil yang diperoleh disajikan pada table grafik dibawah ini.



Gambar 1. Hasil EFA dilihat Menggunakan SPSS 25



Gambar 2. Hasil EFA dilihat Menggunakan JASP 0.95.4

Berdasarkan hasil analisis, terlihat bahwa nilai eigenvalue menunjukkan penurunan yang cukup tajam pada faktor pertama. Hal ini mengindikasikan bahwa instrumen memiliki satu dimensi yang dominan atau bersifat unidimensional. Dengan demikian, sebagian besar variansi data dapat dijelaskan oleh satu konstruk utama yang mendasari instrumen.

Berdasarkan temuan tersebut, hasil uji Exploratory Factor Analysis (EFA) menunjukkan bahwa instrumen penilaian akhir semester pada mata pelajaran Al-Qur'an Hadits memiliki kualitas yang baik. Hal ini ditunjukkan oleh adanya struktur unidimensional, yang mengindikasikan bahwa setiap butir soal dalam instrumen secara konsisten mengukur indikator yang sama sesuai dengan tujuan pengukuran.

Hasil Uji Confirmatory Factor Analysis (CFA)

Analisis Confirmatory Factor Analysis (CFA) dilakukan untuk menguji kesesuaian model konstruk instrumen terhadap data empiris. Pada tahap ini, variabel-variabel pengamatan dianalisis berdasarkan konstruk yang telah ditentukan sebelumnya, sehingga memungkinkan evaluasi kesesuaian model secara lebih terarah. Dalam pelaksanaannya, analisis CFA dilakukan dengan bantuan perangkat lunak JASP. Untuk mempermudah proses analisis, item-item instrumen dikelompokkan ke dalam beberapa kode, yaitu S, T, dan U. Pengelompokan ini bertujuan untuk mengorganisasi data serta memudahkan interpretasi hasil analisis.

Hasil pengujian CFA selanjutnya disajikan dalam bentuk tabel dan gambar yang memuat berbagai indikator kesesuaian model. Indikator tersebut digunakan untuk mengevaluasi sejauh

mana model yang dibangun mampu merepresentasikan struktur konstruk instrumen berdasarkan data empiris yang diperoleh.

Tabel 2. Hasil Chi-Square Test, Indeks Fit, dan RMSEA

Model	χ^2	df	p
Baseline model	2603.666	1225	
Factor model	1677.653	1172	< .001

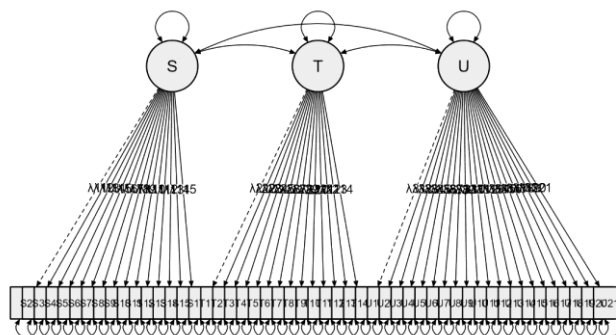
Note. The estimator is ML. The test statistic is standard. The standard error method is standard.

Fit indices

Index	Value
Comparative Fit Index (CFI)	0.633
Tucker-Lewis Index (TLI)	0.617
Bentler-Bonett Non-normed Fit Index (NNFI)	0.617
Bentler-Bonett Normed Fit Index (NFI)	0.356
Parsimony Normed Fit Index (PNFI)	0.340
Bollen's Relative Fit Index (RFI)	0.327
Bollen's Incremental Fit Index (IFI)	0.647
Relative Noncentrality Index (RNI)	0.633

Other fit measures

Metric	Value
Root mean square error of approximation (RMSEA)	0.041
RMSEA 90% CI lower bound	0.037
RMSEA 90% CI upper bound	0.046
RMSEA p-value	1.000
Standardized root mean square residual (SRMR)	0.065
Hoelter's critical N ($\alpha = .05$)	189.923
Hoelter's critical N ($\alpha = .01$)	195.172
Goodness of fit index (GFI)	0.795
McDonald fit index (MFI)	0.368
Expected cross validation index (ECVI)	7.445



Gambar 3. Hasil CFA dilihat Menggunakan JASP 0.95.4

Berdasarkan hasil analisis, diperoleh nilai Chi-square sebesar 1677,653 dengan derajat kebebasan (df) sebesar 1172 dan tingkat signifikansi < 0,001. Selain itu, nilai Root Mean Square Error

of Approximation (RMSEA) sebesar 0,041 menunjukkan bahwa model memiliki tingkat kesesuaian yang cukup baik secara umum. *Other fit measures*

Metric	Value	
Root mean square error of approximation (RMSEA)	0.020	Namun demikian, nilai indeks kesesuaian model khususnya nilai fit instrumen, masih pada angka 0,633, yang menunjukkan bahwa belum sepenuhnya dengan data. Nilai tersebut mengindikasikan bahwa instrumen masih memerlukan penyesuaian untuk mencapai tingkat kesesuaian model yang optimal.
RMSEA 90% CI lower bound	0.010	
RMSEA 90% CI upper bound	0.026	
RMSEA p-value	1.000	
Standardized root mean square residual (SRMR)	0.056	
Hoelter's critical N ($\alpha = .05$)	247.848	
Hoelter's critical N ($\alpha = .01$)	254.846	
Goodness of fit index (GFI)	0.848	
McDonald fit index (MFI)	0.806	
Expected cross validation index (ECVI)	6.064	

kesesuaian model yang optimal.

Oleh karena itu, dilakukan modifikasi model dengan mempertimbangkan hubungan antar item dalam instrumen. Proses modifikasi dilakukan menggunakan perangkat lunak JASP, yang menyediakan informasi dan rekomendasi untuk perbaikan model berdasarkan hasil analisis data. Melalui proses ini, model disesuaikan dengan menghubungkan item-item tertentu yang memiliki keterkaitan, sehingga diharapkan dapat meningkatkan kesesuaian antara model dan data empiris. Hasil analisis setelah dilakukan modifikasi model selanjutnya disajikan pada tabel berikut.

Tabel 3. Hasil Modifikasi Model

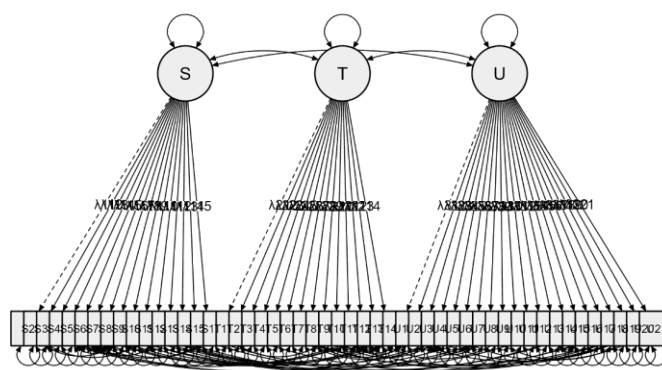
Model Fit

<i>Chi-square test</i>			
Model	X ²	Df	p
Baseline model	2603.666	1225	
Factor model	1234.152	1125	.012

Note. The estimator is ML. The test statistic is standard. The standard error method is standard.

Additional fit measures

<i>Fit indices</i>	
Index	Value
Comparative Fit Index (CFI)	0.921
Tucker-Lewis Index (TLI)	0.914
Bentler-Bonett Non-normed Fit Index (NNFI)	0.914
Bentler-Bonett Normed Fit Index (NFI)	0.526
Parsimony Normed Fit Index (PNFI)	0.483
Bollen's Relative Fit Index (RFI)	0.484
Bollen's Incremental Fit Index (IFI)	0.926
Relative Noncentrality Index (RNI)	0.921



Gambar 4. Hasil Modifikasi Model dilihat Menggunakan JASP 0.95.4

Dari data di atas dapat dilihat bahwasannya nilai Chi-Square adalah 1234,152 dengan nilai df 1125 dan taraf signifikan berada pada angkat 0,12. Nilai RMSEA pada instrumen setelah dimodifikasi berada pada angka 0,020 dan nilai fit instrumen berada pada angka 0,921. Dari hasil tersebut dapat kita lihat bahwasannya setelah dilakukan modifikasi model kualitas konstruk instrumen sudah baik dan bisa digunakan untuk melaksanakan penilaian.

Hasil Uji Reliabilitas Instrumen

Setelah dilakukan pengujian validitas melalui analisis EFA dan CFA, tahap selanjutnya adalah pengujian reliabilitas instrumen untuk menilai tingkat konsistensi internal butir soal. Uji reliabilitas bertujuan untuk mengetahui sejauh mana instrumen mampu memberikan hasil pengukuran yang stabil dan konsisten. Dalam penelitian ini, reliabilitas instrumen dianalisis menggunakan koefisien Cronbach's Alpha. Instrumen dinyatakan reliabel apabila nilai koefisien berada pada kisaran $\geq 0,70$. Perhitungan reliabilitas dilakukan dengan bantuan perangkat lunak SPSS berdasarkan data hasil jawaban peserta didik. Hasil pengujian reliabilitas instrumen penilaian akhir semester disajikan pada tabel berikut.

Tabel 4. Reliabilitas Instrumen Dilihat Menggunakan SPSS 25

RELIABILITY STATISTICS	
CRONBACH'S ALPHA	N OF ITEMS
,791	50

Berdasarkan Tabel 4, nilai Cronbach's Alpha sebesar 0,791 menunjukkan bahwa instrumen memiliki tingkat reliabilitas yang baik. Hal ini mengindikasikan bahwa instrumen memiliki konsistensi internal yang memadai dan dapat digunakan sebagai alat ukur dalam evaluasi pembelajaran.

Discussion

Hasil penelitian ini menunjukkan bahwa kualitas instrumen penilaian akhir semester pada mata pelajaran Al-Qur'an Hadits secara umum telah memenuhi kriteria pengukuran yang baik, meskipun pada tahap awal masih memerlukan penyempurnaan model. Temuan ini memperkuat pentingnya evaluasi instrumen sebagai bagian integral dalam proses penilaian pembelajaran, sebagaimana evaluasi tidak hanya berfungsi mengukur capaian belajar, tetapi juga memastikan akurasi dalam pengambilan keputusan pendidikan (Laila et al., 2024; Muryadi, 2017).

Berdasarkan hasil analisis Exploratory Factor Analysis (EFA), instrumen menunjukkan kecenderungan unidimensional, yang ditandai dengan adanya satu faktor dominan dalam struktur data. Hal ini mengindikasikan bahwa butir-butir soal yang digunakan telah mengarah pada pengukuran konstruk yang relatif homogen. Unidimensionalitas ini menjadi aspek penting dalam pengukuran berbasis teori respons butir, karena memastikan bahwa instrumen benar-benar mengukur satu kemampuan utama secara konsisten (Retnawati, 2016). Temuan ini sejalan dengan pendekatan analisis faktor yang menekankan pentingnya struktur konstruk yang jelas dalam menghasilkan instrumen yang valid (Stapleton, 1997; Putri & Febrilia, 2024).

Namun demikian, hasil Confirmatory Factor Analysis (CFA) pada tahap awal menunjukkan bahwa model belum sepenuhnya sesuai dengan data empiris. Hal ini terlihat dari beberapa indeks kesesuaian model yang masih berada di bawah kriteria ideal, meskipun nilai RMSEA menunjukkan kecocokan yang cukup baik. Kondisi ini mengindikasikan bahwa meskipun struktur dasar instrumen telah terbentuk, masih terdapat hubungan antar item yang belum terakomodasi secara optimal dalam model konstruk. Temuan ini menegaskan bahwa validitas konstruk tidak hanya bergantung pada hasil eksplorasi faktor, tetapi juga pada kesesuaian model secara keseluruhan terhadap data empiris (Kline, 2023).

Proses modifikasi model yang dilakukan dalam penelitian ini terbukti mampu meningkatkan kualitas kesesuaian model secara signifikan. Setelah dilakukan penyesuaian, nilai indeks fit seperti CFI dan TLI meningkat hingga mencapai kategori baik, serta nilai RMSEA menunjukkan tingkat kesesuaian yang sangat baik. Hal ini menunjukkan bahwa modifikasi model merupakan langkah penting dalam analisis CFA untuk memperoleh model yang lebih representatif. Proses ini juga sejalan dengan praktik dalam analisis model struktural yang memungkinkan penyesuaian berdasarkan hubungan empiris antar item guna meningkatkan kesesuaian model (Yamin & Kurniawan, 2009).

Dari perspektif Model Rasch, temuan penelitian ini menunjukkan bahwa analisis kualitas instrumen tidak cukup hanya dilihat dari struktur konstruk, tetapi juga perlu mempertimbangkan karakteristik butir dan kesesuaian model pengukuran. Nilai fit instrumen yang belum optimal pada tahap awal menunjukkan adanya butir yang tidak sepenuhnya sesuai dengan model, sehingga memerlukan perbaikan. Model Rasch memberikan keunggulan dalam mengidentifikasi ketidaksesuaian tersebut secara lebih objektif dan independen terhadap sampel, sehingga memungkinkan evaluasi instrumen yang lebih akurat (Boone, 2016; Yudha, 2023; Scoulas et al., 2021). Hal ini memperkuat argumen bahwa pendekatan berbasis Item Response Theory lebih unggul dibandingkan teori tes klasik dalam menghasilkan pengukuran yang objektif dan stabil (Chan et al., 2021; Pratama, 2020).

Selain validitas, hasil analisis reliabilitas menunjukkan bahwa instrumen memiliki konsistensi internal yang baik dengan nilai Cronbach's Alpha sebesar 0,791. Nilai ini menunjukkan bahwa instrumen mampu memberikan hasil pengukuran yang stabil dan dapat diandalkan. Reliabilitas yang baik menjadi indikator penting dalam kualitas instrumen, karena memastikan bahwa hasil evaluasi tidak bersifat fluktuatif dan dapat digunakan sebagai dasar dalam pengambilan keputusan pembelajaran (Al Hakim et al., 2021; Devi et al., 2022). Dengan demikian, instrumen yang digunakan dalam penelitian ini telah memenuhi standar minimal reliabilitas dalam evaluasi pendidikan.

Secara keseluruhan, temuan penelitian ini menguatkan bahwa integrasi antara Model Rasch dengan analisis faktor (EFA dan CFA) memberikan pendekatan yang lebih komprehensif dalam mengevaluasi kualitas instrumen. Analisis EFA berperan dalam mengidentifikasi struktur dimensi awal, CFA menguji kesesuaian model konstruk, sementara Model Rasch memberikan informasi mengenai karakteristik butir dan kemampuan responden secara simultan. Kombinasi ketiga pendekatan ini mampu menjawab kesenjangan penelitian yang sebelumnya diidentifikasi, yaitu keterbatasan analisis yang hanya dilakukan secara parsial (Azzahra et al., 2024; Syafii et al., 2021).

Implikasi dari penelitian ini menunjukkan bahwa pengembangan instrumen evaluasi pembelajaran, khususnya dalam konteks pendidikan agama Islam, perlu dilakukan secara sistematis dan berbasis data empiris. Instrumen yang tidak melalui proses analisis yang komprehensif berpotensi menghasilkan pengukuran yang bias dan tidak akurat, sehingga dapat mempengaruhi kualitas pembelajaran secara keseluruhan. Oleh karena itu, penerapan pendekatan analisis terpadu seperti yang digunakan dalam penelitian ini dapat menjadi alternatif yang efektif dalam meningkatkan kualitas instrumen evaluasi serta mendukung praktik penilaian yang lebih objektif dan berbasis bukti.

CONCLUSION

Penelitian ini menunjukkan bahwa kualitas instrumen penilaian akhir semester pada mata pelajaran Al-Qur'an Hadits kelas VII di MTsN 2 Kota Bukittinggi secara umum berada pada kategori baik berdasarkan analisis yang terintegrasi. Hasil uji Exploratory Factor Analysis (EFA) mengindikasikan bahwa instrumen bersifat unidimensional, yang ditunjukkan oleh dominasi satu

faktor utama dalam struktur data. Pada tahap Confirmatory Factor Analysis (CFA), model awal menunjukkan kesesuaian yang belum optimal, khususnya pada nilai indeks fit yang masih rendah, sehingga diperlukan modifikasi model. Setelah dilakukan penyesuaian, hasil analisis menunjukkan peningkatan signifikan pada indeks kesesuaian model, yang mengindikasikan bahwa konstruk instrumen telah sesuai dengan data empiris. Selain itu, hasil uji reliabilitas menunjukkan nilai Cronbach's Alpha sebesar 0,791, yang menandakan bahwa instrumen memiliki konsistensi internal yang baik dan dapat diandalkan dalam pengukuran. Dengan demikian, penelitian ini menegaskan bahwa penggunaan pendekatan terpadu antara Model Rasch, EFA, dan CFA mampu memberikan evaluasi kualitas instrumen yang lebih komprehensif dan akurat. Berdasarkan temuan tersebut, disarankan agar pengembangan instrumen evaluasi pembelajaran dilakukan secara sistematis dengan memanfaatkan analisis empiris yang terintegrasi, serta tidak hanya berfokus pada ranah kognitif, tetapi juga mencakup aspek afektif dan psikomotorik agar pengukuran hasil belajar menjadi lebih holistik.

ACKNOWLEDGMENT

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada MTs N 2 Kota Bukittinggi yang telah memberikan kontribusi signifikan dalam penelitian ini. Kontribusi tersebut meliputi penyediaan waktu, tempat, serta akses data penelitian yang meliputi naskah soal, kunci jawaban, dan lembar jawaban siswa kelas VII pada mata pelajaran Al-Qur'an Hadits. Tanpa dukungan dan kerja sama yang baik dari pihak madrasah, penelitian ini tidak dapat terlaksana dengan lancar dan dipublikasikan. Semoga kontribusi ini menjadi amal jariyah dan bermanfaat bagi kemajuan pendidikan, khususnya dalam pengembangan evaluasi pembelajaran di MTsN 2 Kota Bukittinggi.

AUTHOR CONTRIBUTIONS STATEMENT

UR bertanggung jawab atas konseptualisasi penelitian, perancangan metodologi, supervisi, serta penyusunan dan revisi draf manuskrip. RF berperan dalam analisis data menggunakan model Rasch, interpretasi hasil penelitian, serta kontribusi dalam penulisan bagian metodologi dan pembahasan. SR bertanggung jawab atas pengolahan data mentah, dokumentasi, serta tabulasi hasil penelitian. Sementara itu, RI berperan dalam melakukan review kritis terhadap manuskrip, validasi hasil analisis, serta penyuntingan akhir naskah sebelum dipublikasikan.

REFERENCES

- 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>
- Ariesa, Y. (2019). Faktor-faktor yang mempengaruhi ketahanan pangan dengan menggunakan analisis faktor konfirmatori: Yeni Ariesa; Rafida Khairani. *Jurnal Ilmiah Abdi Ilmu*, 12(1), 8-18.
- Arikunto, S. (2021). *Prosedur penelitian suatu pendekatan praktik*. Rineka Cipta.
- Azzahra, N. S., Sumarni, S., & Putranta, H. (2024). Analisis validitas dan realibilitas kualitas soal pilihan ganda Asesmen Kompetensi Minimum (AKM) mata pelajaran pendidikan agama islam menggunakan pendekatan model rasch. *QuranicEdu: Journal of Islamic Education*, 4(1), 85-94. <https://doi.org/10.37252/quranicedu.v4i1.681>
- Boone, W. J. (2016). Rasch analysis for instrument development: why, when, and how?. *CBE—Life Sciences Education*, 15(4), rm4. <https://doi.org/10.1187/cbe.16-04-0148>
- Chan, S. W., Looi, C. K., & Sumintono, B. (2021). Assessing computational thinking abilities among Singapore secondary students: a Rasch model measurement analysis. *Journal of Computers in Education*, 8(2), 213-236. <https://doi.org/10.1007/s40692-020-00177-2>
- Devi, M. Y., Hidayanthi, R., & Fitria, Y. (2022). Model-model evaluasi pendidikan dan model sepuluh langkah dalam penilaian. *Jurnal Basicedu*, 6(1), 675-683. <https://doi.org/10.31004/basicedu.v6i1.1934>
- Ghazali, N., & Nordin, M. S. (2019). Measuring meaningful learning experience: Confirmatory factor analysis. *International Journal of Innovation, Creativity and Change*, 9(12), 283-296.

- Hambleton, R. K., & Swaminathan, H. (2013). *Item response theory: Principles and applications*. Springer Science & Business Media. <https://doi.org/10.1007/978-94-017-1988-9>
- Heri, L., Rusilowati, A., & Raharjo, T. J. (2017). Pengembangan instrumen penilaian psikomotor senam lantai dalam pembelajaran penjasorkes pada siswa sekolah dasar. *Journal of Research and Educational Research Evaluation*, 6(1), 19-29. <https://doi.org/10.15294/jrer.v6i1.16204>
- Khumaedi, M. (2012). Reliabilitas instrumen penelitian pendidikan. *Jurnal Pendidikan Teknik Mesin*, 12(1), 25-30. <https://doi.org/10.15294/jptm.v12i1.5273>
- Kline, R. B. (2023). *Principles and practice of structural equation modeling (4th ed.)*. Guilford Press.
- Laila, L., Nabila, A., & Widyanti, E. (2024). Konsep dasar evaluasi pembelajaran. *Jurnal Manajemen Dan Pendidikan Agama Islam*, 2(5), 252-262. <https://doi.org/10.61132/jmpai.v2i5.536>
- Linn, R. L. (2006). *Educational measurement (4th ed.)*. Praeger Publishers.
- Muryadi, A. D. (2017). Model evaluasi program dalam penelitian evaluasi. *Jurnal Ilmiah Penjas (Penelitian, Pendidikan Dan Pengajaran)*, 3(1).
- Natanael, Y., & Novanto, Y. (2020). Pengujian model pengukuran congeneric, tau-Equivalent dan parallel pada satisfaction with life scale (SWLS). *Psymphatic: Jurnal Ilmiah Psikologi*, 7(2), 285-298. <https://doi.org/10.15575/psy.v7i2.6405>
- Prasetyo, Y. T., Senoro, D. B., German, J. D., Robielos, R. A. C., & Ney, F. P. (2020). Confirmatory factor analysis of vulnerability to natural hazards: A household Vulnerability Assessment in Marinduque Island, Philippines. *International Journal of Disaster Risk Reduction*, 50, 101831. <https://doi.org/10.1016/j.ijdrr.2020.101831>
- Pratama, D. (2020). Analisis kualitas tes buatan guru melalui pendekatan item response theory (IRT) model rasch. *Tarbawy: Jurnal Pendidikan Islam*, 7(1), 61-70. <https://doi.org/10.32923/tarbawy.v7i1.1187>
- Putri, R. D. P., & Febrilia, Y. (2024). Confirmatory Factor Analysis (CFA) on the quality of critical thinking instruments, motivation, and learning achievement of learners. *PENDAGOGIA: Jurnal Pendidikan Dasar*, 4(2), 73-84.
- Retnawati, H. (2016). *Analisis kuantitatif instrumen penelitian (panduan peneliti, mahasiswa, dan psikometrian)*. Parama publishing.
- Santoso, P. H., Istiyono, E., Haryanto, & Hidayatulloh, W. (2022). Thematic analysis of Indonesian physics education research literature using machine learning. *Data*, 7(11), 147. <https://doi.org/10.3390/data7110147>
- Scoulas, J. M., Dunya, B. A., & De Groote, S. L. (2021). Validating students' library experience survey using rasch model. *Library & Information Science Research*, 43(1), 101071. <https://doi.org/10.1016/j.lisr.2021.101071>
- Stapleton, C. D. (1997, January). Basic concepts in exploratory factor analysis (EFA) as a tool to evaluate score validity: A right-brained approach. Paper presented at the Annual Meeting of the Southeast Educational Research Association, Austin, TX, United States.
- Sugiono, S., Noerdjanah, N., & Wahyu, A. (2020). Uji validitas dan reliabilitas alat ukur SG posture evaluation. *Jurnal Keterampilan Fisik*, 5(1), 55-61. <https://doi.org/10.37341/jkf.v5i1.167>
- Syafii, A., Haryanto, H., Ahmad, I. F., & Fauziah, A. (2021). Analysis of items with Item Response Theory (IRT) approach on final assessment for al-quran hadith subjects. *Jurnal Pendidikan Agama Islam*, 18(1), 167-194. <https://doi.org/10.14421/jpai.2021.181-09>
- Unaradjan, D. D. (2019). *Metode penelitian kuantitatif*. Penerbit Unika Atma Jaya Jakarta.
- Yamin, S., & Kurniawan, H. (2009). SPSS complete: Teknik analisis statistik terlengkap dengan software SPSS. *Jakarta: Salemba Infotek*, 328.
- Yudha, R. P. (2023). Higher order thinking skills (HOTS) test instrument: Validity and reliability analysis with the rasch model. *EduMa: Mathematics education learning and teaching*, 12(1), 21-38. <https://doi.org/10.24235/eduma.v12i1.9468>