

Analisis Model Rasch dalam Pengembangan Kuesioner Validasi Teoritik Butir Soal Pilihan Ganda Instrumen Tes Diagnostik

Slamet Maulana^{1*}, Elvara Norma Aroyandini²

¹ SMA Labschool Cibubur

² Program Studi Pendidikan Guru dan Sekolah Dasar, Universitas Terbuka

*Surel Penulis Koresponden: slametmaulana1996@gmail.com

Abstrak

Pengembangan instrumen tes diagnostik saat ini merupakan salah satu solusi untuk menganalisis masalah miskonsepsi. Dengan demikian, diperlukan instrumen yang dapat digunakan untuk secara teoritis memvalidasi instrumen tes diagnostik yang akan dikembangkan. Model Rasch merupakan salah satu model analisis yang dapat digunakan untuk menentukan kelayakan suatu instrumen. Penelitian ini bertujuan untuk mengembangkan kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik melalui implementasi Model Rasch. Penelitian ini merupakan penelitian dan pengembangan dengan model ADDIE. Penelitian ini telah menghasilkan suatu instrumen validasi teoritis berbentuk kuesioner untuk memvalidasi butir soal berbentuk pilihan ganda pada suatu instrument tes diagnostik. Instrumen tes diagnostik yang telah dikembangkan terbagi menjadi tiga aspek, yaitu aspek materi sebanyak delapan butir, aspek konstruksi soal sebanyak 14 butir, dan aspek bahasa sebanyak enam butir. Hasil analisis Model Rasch menyatakan bahwa kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik valid dari segi konstruk dan konstruk, serta reliabel. Penelitian ini menghasilkan kesimpulan bahwa produk pendidikan berupa kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik dinyatakan layak dan dapat digunakan.

Kata Kunci: Model Rasch, Kuesioner Validasi Teoritik, Instrumen Tes Diagnostik.

Abstract

The development of diagnostic test instruments is currently one of the solutions to analyze the problem of misconceptions. Thus, an instrument that can theoretically validate the diagnostic test instrument to be developed is needed. The Rasch model is one of the analytical models that can be used to determine the feasibility of an instrument. This study aims to create a theoretical validation questionnaire of multiple-choice items on diagnostic test instruments by implementing the Rasch Model. This research is a research and development with the ADDIE model. This research has produced a theoretical validation instrument in the form of a questionnaire to validate multiple-choice question items on a diagnostic test instrument. The diagnostic test instrument that has been developed is divided into three aspects: the material aspect, which has as many as eight items; the question construction aspect, which has as many as 14 items; and the language aspect, which has as many as six items. The results of the Rasch Model analysis state that the theoretical validation questionnaire of multiple-choice question items on diagnostic test instruments is valid in terms of construct and



construct and is reliable. This study concluded that educational products in the form of theoretical validation questionnaires of multiple-choice question items on diagnostic test instruments were feasible and could be used.

Keywords: *Rasch Model, Theoretical Validation Questionnaire, Diagnostic Test Instrument.*

A. PENDAHULUAN

Kesalahpahaman masih menjadi momok masalah dalam dunia pendidikan. Alasannya adalah bahwa pemahaman konsep yang kuat memainkan peran penting dalam membantu peserta didik mengembangkan basis pengetahuan dan struktur mereka, menerapkan konsep yang benar untuk pemecahan masalah, dan dengan demikian mengembangkan keterampilan dan kompetensi dalam profesi mereka.

Pengembangan instrumen tes diagnostik saat ini merupakan salah satu solusi untuk menganalisis masalah miskonsepsi. Melalui serangkaian tes diagnostik ini, kesalahpahaman yang dialami oleh siswa dapat dengan mudah diidentifikasi dan dianalisis. Tes diagnostik yang baik dapat memberikan gambaran yang akurat tentang kesalahpahaman yang dialami siswa berdasarkan informasi yang dia buat kesalahan (Wilantika *et al.*, 2018).

Namun, apakah serangkaian tes diagnostik yang dikembangkan seperti itu dapat langsung digunakan untuk umum. Tentu saja, serangkaian tes diagnostik seperti itu harus memiliki kekuatan yang tepat. Sehingga ketika digunakan, tes diagnostik tersebut mampu mengukur apa yang seharusnya diukur. Akibatnya, hasil analisis yang diberikan oleh serangkaian tes diagnostik dapat dipercaya. Dengan demikian, diperlukan instrumen yang dapat digunakan untuk secara teoritis memvalidasi instrumen tes diagnostik yang akan dikembangkan.

Model Rasch merupakan salah satu model analisis yang dapat digunakan untuk menentukan kelayakan suatu instrumen. Model Rasch telah digunakan dalam analisis berbagai instrumen penilaian dalam memberikan informasi rinci tentang beberapa aspek kualitas tes (Ling *et al.*, 2014). Model Rasch dapat digunakan dalam menentukan indeks reliabilitas untuk menganalisis item ke setiap level, reliabilitas responden, serta dimensi dan mendeteksi adanya bias dari item pada instrumen (Suryani, 2018).

Model Analisis Rasch merupakan pendekatan dinamis untuk mencapai validitas konstruk (Semakula *et al.*, 2017). Analisis Model Rasch mendukung dalam pengembangan dan validasi pengukuran invarian dalam memberikan bukti empiris dan wawasan untuk sifat psikometrik penting dari tes dan kuesioner (Sjaastad, 2014). Analisis Model Rasch mendukung dalam pengembangan dan validasi pengukuran invarian dalam memberikan bukti

empiris dan wawasan untuk sifat psikometrik penting dari tes dan kuesioner (Fulmer *et al.*, 2016). Model Rasch juga mampu menentukan kemampuan siswa, dimana probabilitas seorang siswa menjawab suatu barang dengan benar akan meningkat secara monoton dengan kemampuan yang dimiliki siswa (Xiao *et al.*, 2018).

Rasch Model tidak hanya digunakan untuk menganalisis item instrumen, tetapi juga dapat menganalisis peserta didik (Matore *et al.*, 2018; Zamri & Nordin, 2015). Model Rasch juga mampu menentukan kemampuan siswa, dimana probabilitas seorang siswa menjawab suatu barang dengan benar akan meningkat secara monoton dengan kemampuan yang dimiliki siswa (Islam *et al.*, 2020; Xiao *et al.*, 2018).

Berdasarkan uraian di atas, penelitian ini bertujuan untuk mengembangkan kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik melalui analisis Model Rasch.

B. METODE PENELITIAN

Penelitian ini merupakan penelitian dan pengembangan dengan metode penelitian adalah kuantitatif analisis. Model penelitian dan pengembangan yang digunakan adalah model ADDIE. Penelitian ini terdiri dari lima tahap, yaitu Analyze, Design, Develop, Implement, dan Evaluation.

Pada tahap analisis, penelitian diawali dengan melakukan studi pustaka dan observasi ke lapangan. Hasil studi pustaka menunjukkan bahwa miskonsepsi masih menjadi salah satu masalah pendidikan. Kesalahpahaman tersebut dapat diukur melalui serangkaian tes diagnostik. Namun, serangkaian tes diagnostik memerlukan validasi terlebih dahulu sebelum diterapkan secara umum. Hal ini juga didukung oleh pengamatan bahwa sejumlah guru sudah memiliki seperangkat tes diagnostik, namun belum mengetahui validitas alat tes tersebut. Sehingga pada tahap ini diperoleh tujuan untuk mengembangkan instrumen validasi teoritis yang valid dalam mengukur validitas suatu instrumen tes diagnostik.

Pada tahap desain, *blueprint* instrumen validasi teoritis mulai dikembangkan. Instrumen validasi teoritis yang akan dikembangkan mengacu pada kesesuaian materi dan fungsi tes diagnostik dalam menganalisis miskonsepsi pada siswa. Instrumen validasi teoritis yang akan dikembangkan menggunakan skala likert dengan lima skala untuk pengisian. Adapun rencana instrument validasi teoritis tersebut akan dikembangkan dalam bentuk kuesioner

Pada tahap pengembangan, *blueprint* kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik yang telah selesai mulai dikembangkan dan direalisasikan. Instrumen validasi teoritis yang

dikembangkan memiliki 28 butir pernyataan dengan skala likert 1-5. Kriteria skala likert yang digunakan adalah 1 untuk sangat tidak setuju, 2 untuk tidak setuju, 3 untuk ragu-ragu atau netral, 4 untuk setuju, dan 5 untuk sangat setuju. Kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik yang telah dikembangkan kemudian divalidasi terlebih dahulu dengan menghitung nilai *Content Validity Ratio (CVR)* dan *Content Validity Index (CVI)*. Pada tahap ini, CVR digunakan untuk mengetahui validitas isi setiap butir pernyataan pada kuesioner. Adapun nilai CVI digunakan untuk mengetahui validitas instrument kuesioner secara keseluruhan.

Pada tahap implementasi, kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik yang telah dikembangkan dan telah dinyatakan valid melalui skor CVR dan CVI diimplementasikan di lapangan. Kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik yang telah dikembangkan kemudian digunakan untuk menilai kualitas suatu alat uji diagnostik yang telah dikembangkan pada penelitian sebelumnya (Susanti *et al.*, 2021). Hasil pengisian kemudian dikumpulkan dan dianalisis melalui model Rasch. Melalui implementasi model Rasch, validasi empiris dilakukan untuk mengetahui validitas konstruk dari kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik yang telah dikembangkan.

Pada tahap evaluasi, setiap tahapan yang telah dilakukan dievaluasi. Hal ini dilakukan dalam menjaga kualitas pelaksanaan penelitiannya. Sehingga setiap tahapan penelitian yang telah dilakukan dievaluasi terlebih dahulu sebelum dilanjutkan ke tahap berikutnya. Adapun evaluasi akhir dilakukan untuk memperoleh kesimpulan akhir pada penelitian ini.

C. HASIL DAN PEMBAHASAN

Penelitian ini telah menghasilkan suatu instrumen validasi teoritis berbentuk kuesioner untuk memvalidasi butir soal berbentuk pilihan ganda pada suatu instrument tes diagnostik. Kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik yang telah dikembangkan memiliki 28 item pernyataan dengan skala likert lima tingkat. Kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik yang telah dikembangkan terbagi menjadi tiga aspek, yaitu aspek materi sebanyak 8 butir, aspek konstruksi soal sebanyak 14 butir, dan aspek bahasa sebanyak 6 butir. Adapun daftar butir pernyataan yang telah dikembangkan dapat dilihat pada tabel 1.

Tabel 1. Butir Pernyataan Kuesioner Validasi Teoritik Butir Soal Pilihan

Ganda pada Instrumen Tes Diagnostik.

| No | Aspek | Butir Pernyataan |
|----|-----------------|---|
| 1 | Materi | Setiap butir soal mengukur tujuan pembelajaran yang diharapkan. |
| 2 | | Setiap butir soal sesuai dengan indikator pada kisi-kisi. |
| 3 | | Pertanyaan dan jawaban pada setiap butir soal berkesinambungan. |
| 4 | | Stimulus pada setiap butir soal berkesinambungan. |
| 5 | | Setiap butir soal bersifat objektif. |
| 6 | | Pertanyaan pada setiap butir soal jelas. |
| 7 | | Jawaban pada setiap butir soal jelas. |
| 8 | | Materi yang diujikan pada setiap butir soal benar. |
| 9 | Konstruksi Soal | Pokok pertanyaan di dalam setiap butir soal singkat. |
| 10 | | Pokok pertanyaan di dalam setiap butir soal jelas. |
| 11 | | Pokok pertanyaan di dalam setiap butir soal tegas. |
| 12 | | Isi pokok pertanyaan di dalam setiap butir soal hanya mengandung pernyataan yang diperlukan. |
| 13 | | Pokok pertanyaan di dalam setiap butir soal tidak memberikan informasi yang mengarahkan kepada jawaban yang benar. |
| 14 | | Pokok pertanyaan di dalam setiap butir soal tidak mengandung pernyataan yang bersifat negatif ganda. |
| 15 | | Pilihan jawaban di dalam setiap butir soal hanya mengandung pernyataan yang diperlukan. |
| 16 | | Pilihan jawaban di dalam setiap butir soal bersifat homogen. |
| 17 | | Pilihan jawaban di dalam setiap butir soal tidak menggunakan pilihan “semua jawaban benar” atau “semua jawaban salah”. |
| 18 | | Pilihan jawaban di dalam setiap butir soal disusun berdasarkan urutan angka atau urutan suatu kejadian. |
| 19 | | Butir jawaban di dalam setiap butir soal tidak memiliki ketergantungan dengan butir jawaban dari butir soal sebelumnya. |
| 20 | | Stimulus (gambar/suara/animasi/video) di dalam setiap butir soal dapat jelas terlihat. |

| No | Aspek | Butir Pernyataan |
|----|--------|--|
| 21 | | Stimulus (gambar/suara/animasi/video) di dalam setiap butir soal berfungsi. |
| 22 | | Ukuran huruf di dalam setiap butir soal dapat terbaca. |
| 23 | Bahasa | Bahasa yang digunakan dalam setiap butir soal menggunakan kaidah Bahasa Indonesia yang baik dan benar. |
| 24 | | Bahasa yang digunakan dalam setiap butir soal singkat. |
| 25 | | Bahasa yang digunakan dalam setiap butir soal jelas. |
| 26 | | Bahasa yang digunakan dalam setiap butir soal komunikatif. |
| 27 | | Tidak terjadi pengulangan kata tidak bermakna di dalam setiap butir soal. |
| 28 | | Setiap butir soal tidak menggunakan bahasa lokal atau setempat. |

1. CVR dan CVI

CVR dihitung untuk mengetahui validitas konten setiap butir pernyataan pada suatu instrumen yang telah dikembangkan. Sedangkan nilai CVI digunakan untuk mengetahui validitas suatu instrumen secara keseluruhan. Perhitungan nilai CVR dan CVI kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik yang telah dikembangkan melibatkan sebelas validator untuk memberikan penilaian apakah setiap item dalam instrumen valid atau tidak. Oleh karena itu, nilai CVR minimum yang harus terpenuhi agar suatu butir pernyataan pada kuesioner dinyatakan valid secara konten adalah 0,59. Adapun nilai CVR untuk setiap butir dapat dilihat pada tabel 2.

Tabel 2. Nilai CVR Butir Pernyataan Instrumen Validasi Teoritis

| No Butir | Nilai CVR | No Butir | Nilai CVR |
|----------|-----------|----------|-----------|
| 1 | 0,64 | 15 | 0,82 |
| 2 | 1,00 | 16 | 0,82 |
| 3 | 1,00 | 17 | 0,64 |

| No Butir | Nilai CVR | No Butir | Nilai CVR |
|----------|-----------|----------|-----------|
| 4 | 0,64 | 18 | 1,00 |
| 5 | 0,64 | 19 | 0,82 |
| 6 | 0,64 | 20 | 0,64 |
| 7 | 0,82 | 21 | 1,00 |
| 8 | 0,82 | 22 | 0,64 |
| 9 | 0,82 | 23 | 0,64 |
| 10 | 0,82 | 24 | 0,82 |
| 11 | 0,82 | 25 | 1,00 |
| 12 | 1,00 | 26 | 0,82 |
| 13 | 0,64 | 27 | 0,82 |
| 14 | 0,64 | 28 | 0,64 |

Hasil perhitungan CVR setiap butir pernyataan di dalam kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik yang telah dikembangkan yang telah dikembangkan memberikan nilai CVR untuk setiap butir lebih dari 0,59. Hasil ini menginformasikan bahwa setiap butir item pada instrumen validasi teoritis dinyatakan valid secara konten.

Nilai tersebut menjelaskan bahwa setiap butir pernyataan yang telah dikembangkan telah dinyatakan valid secara konten dan dapat digunakan. Hal ini sejalan dengan penelitian sebelumnya tentang pengembangan suatu instrumen yang menyatakan bahwa validitas konten (melalui perhitungan CVR) sangat penting dan krusial (Alizadeh-Siuki *et al.*, 2020; Waltz *et al.*, 2016).

Selanjutnya, hasil perhitungan CVI pada kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik yang telah

dikembangkan memberikan nilai CVI sebesar 0,79. Hasil ini menginformasikan bahwa kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik yang telah dikembangkan dinyatakan valid dengan kriteria sangat baik. Oleh karena itu, kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik yang telah dikembangkan dapat dinyatakan valid baik di setiap item maupun di keseluruhan instrumen.

Setiap butir pernyataan yang dinyatakan valid belum tentu menjadi valid jika telah disatukan menjadi instrumen tes. Jadi, setelah setiap soal dinyatakan valid berdasarkan kriteria minimal nilai CVR, maka dilakukan perhitungan nilai CVI terhadap kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik yang telah dikembangkan. Hasil perhitungan nilai CVI memberikan nilai yang juga melebihi batas minimum. Ini menjelaskan bahwa semua item pernyataan yang telah dikembangkan ketika ditinjau secara keseluruhan sebagai instrumen keseluruhan dinyatakan valid dan dapat digunakan. Validitas konten memastikan bahwa item dan instrumen uji yang telah dirancang dan dikembangkan dengan benar dapat mengukur (Al-Taweel & Awad, 2020).

2. Analisis Model Rasch

Hasil analisis Model Rasch menunjukkan bahwa kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik yang telah dikembangkan memiliki nilai *Raw Variance* sebesar 27,6%. Hal ini menginformasikan bahwa kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik memiliki validitas konstruk yang baik. Ini menginformasikan bahwa kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik benar-benar mengukur apa yang seharusnya diukur.

| Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units | | | | |
|--|---|------------|----------|----------|
| | | Eigenvalue | Observed | Expected |
| Total raw variance in observations | = | 23.2328 | 100.0% | 100.0% |
| Raw variance explained by measures | = | 6.2328 | 26.8% | 27.6% |
| Raw variance explained by persons | = | 2.8292 | 12.2% | 12.5% |
| Raw Variance explained by items | = | 3.4035 | 14.6% | 15.1% |
| Raw unexplained variance (total) | = | 17.0000 | 73.2% | 100.0% |
| Unexplned variance in 1st contrast | = | 2.1704 | 9.3% | 12.7% |
| Unexplned variance in 2nd contrast | = | 1.9423 | 8.4% | 11.5% |
| Unexplned variance in 3rd contrast | = | 0.6884 | 3.0% | 4.1% |
| Unexplned variance in 4th contrast | = | 0.6797 | 2.9% | 3.9% |
| Unexplned variance in 5th contrast | = | 0.4458 | 1.9% | 2.6% |

Gambar 1. Raw variance kuesioner validasi teoritik.

Model Rasch digunakan karena mampu mendefinisikan validitas konstruk butir (Mohamad *et al.*, 2015). Adapun pernyataan yang dianalisis menggunakan Model Rasch dan dinyatakan sesuai, maka hasilnya cukup memberikan informasi bahwa pertanyaan yang dikembangkan valid dan dapat mengukur apa yang harus diukur (Boone, 2016; Planinic *et al.*, 2019).

Raw variance menjadi sifat laten dari instrumen yang diukur (Al Zoubi *et al.*, 2018; Nielsen & Dammeyer, 2019) dalam memverifikasi konstruksi pengukuran instrumen tersebut (Chang *et al.*, 2020; Lo *et al.*, 2015). Validitas tidak hanya menginformasikan tentang validitas konten tes, tetapi juga menginformasikan konsekuensi dari penggunaan skor tes (Mokshein *et al.*, 2019). Ketika menggunakan model Rasch, hasil analisis dapat menetapkan kesesuaian item dengan konstruksi yang diidentifikasi (Yunos *et al.*, 2017) terhadap instrumen secara keseluruhan (Yasin *et al.*, 2015; Yasin *et al.*, 2018).

Hasil analisis Model Rasch lainnya juga menunjukkan bahwa kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik memiliki reliabilitas yang cukup baik dengan nilai reliabilitas sebesar 0,75. Hal ini menginformasikan bahwa butir soal dalam kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik memiliki reliabilitas yang cukup baik.

| | TOTAL SCORE | COUNT | MEASURE | MODEL S.E. | INFIT MNSQ | ZSTD | OUTFIT MNSQ | ZSTD |
|-------------------------|-------------|---------|---------|------------|------------|------|-------------|-------|
| MEAN | 60.7 | 73.0 | .00 | .37 | 1.00 | .09 | .96 | .01 |
| SEM | 1.4 | .0 | .18 | .02 | .02 | .15 | .06 | .20 |
| P.SD | 6.1 | .0 | .76 | .09 | .09 | .64 | .28 | .88 |
| S.SD | 6.2 | .0 | .78 | .10 | .09 | .66 | .29 | .90 |
| MAX. | 70.0 | 73.0 | 1.25 | .60 | 1.21 | 1.96 | 1.80 | 2.12 |
| MIN. | 48.0 | 73.0 | -1.51 | .27 | .86 | -.97 | .48 | -1.25 |
| REAL RMSE | .38 | TRUE SD | .66 | SEPARATION | 1.71 | Item | RELIABILITY | .75 |
| MODEL RMSE | .38 | TRUE SD | .66 | SEPARATION | 1.74 | Item | RELIABILITY | .75 |
| S.E. OF Item MEAN = .18 | | | | | | | | |

Gambar 2. Reliabilitas kuesioner validasi teoritik.

Selain itu dalam analisis Model Rasch, reliabilitas memberikan informasi yang lebih rinci antara sifat-sifat item pada instrumen dan individu yang merespons item tersebut (Bodzin *et al.*, 2020). Reliabilitas juga menginformasikan tentang konsistensi butir pernyataan pada suatu instrumen (Koçak, 2020).

Hasil analisis Model Rasch kemudian dilanjutkan untuk menentukan konsistensi internal setiap butir pernyataan dalam kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik yang telah dikembangkan. Item pernyataan dinyatakan sesuai jika memenuhi salah satu dari tiga kondisi. Ketiga kondisi tersebut adalah nilai *Outfit Mean Square* antara 0,5 hingga 1,5; nilai *Outfit Z-Standard* antara -2,0 hingga 2,0; dan nilai *Point Measure Correlation* antara 0,4 hingga 0,85. Hasil analisis Model Rasch menunjukkan bahwa semua item pernyataan dalam kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik yang telah dikembangkan dinyatakan layak karena memenuhi setidaknya satu dari tiga persyaratan konsistensi internal.

| ENTRY NUMBER | TOTAL SCORE | TOTAL COUNT | JMLE MEASURE | MODEL S. E. | INFIT MNSQ | ZSTD | OUTFIT MNSQ | ZSTD | PTMEASUR-AL CORR. | EXP. | EXACT OBS% | MATCH EXP% | Item | |
|--------------|-------------|-------------|--------------|-------------|------------|------|-------------|-------|-------------------|------|------------|------------|------|-----|
| 10 | 62 | 73 | .00 | .34 | 1.18 | .87 | 1.80 | 2.12 | A | .07 | .29 | 83.3 | 83.3 | S10 |
| 11 | 59 | 73 | .33 | .32 | 1.12 | .76 | 1.28 | 1.09 | B | .20 | .32 | 78.8 | 78.9 | S11 |
| 20 | 48 | 73 | 1.25 | .27 | 1.21 | 1.96 | 1.27 | 1.86 | C | .24 | .41 | 60.6 | 67.5 | S20 |
| 13 | 63 | 73 | -.12 | .36 | .98 | -.02 | 1.16 | .56 | D | .27 | .28 | 84.8 | 84.9 | S13 |
| 9 | 63 | 73 | -.12 | .36 | 1.11 | .53 | 1.12 | .46 | E | .19 | .28 | 84.8 | 84.9 | S9 |
| 1 | 55 | 73 | .70 | .29 | .97 | -.17 | 1.06 | .36 | F | .37 | .36 | 75.8 | 73.7 | S1 |
| 3 | 63 | 73 | -.12 | .36 | 1.04 | .24 | .97 | .04 | G | .26 | .28 | 84.8 | 84.9 | S3 |
| 19 | 55 | 73 | .70 | .29 | 1.03 | .29 | 1.04 | .26 | H | .33 | .36 | 69.7 | 73.7 | S19 |
| 16 | 65 | 73 | -.40 | .39 | 1.02 | .16 | .86 | -.23 | I | .25 | .25 | 87.9 | 87.9 | S16 |
| 4 | 67 | 73 | -.74 | .44 | .99 | .09 | .80 | -.28 | J | .24 | .22 | 90.9 | 90.9 | S4 |
| 5 | 67 | 73 | -.74 | .44 | .99 | .09 | .83 | -.20 | j | .24 | .22 | 90.9 | 90.9 | S5 |
| 8 | 68 | 73 | -.95 | .48 | .99 | .09 | .96 | .10 | i | .21 | .20 | 92.4 | 92.4 | S8 |
| 12 | 58 | 73 | .43 | .31 | .99 | -.01 | .93 | -.21 | h | .34 | .33 | 80.3 | 77.5 | S12 |
| 6 | 70 | 73 | -1.51 | .60 | .98 | .14 | .65 | -.29 | g | .21 | .15 | 95.5 | 95.5 | S6 |
| 2 | 59 | 73 | .33 | .32 | .97 | -.12 | .82 | -.67 | f | .37 | .32 | 78.8 | 78.9 | S2 |
| 7 | 70 | 73 | -1.51 | .60 | .92 | .02 | .48 | -.62 | e | .27 | .15 | 95.5 | 95.5 | S7 |
| 18 | 49 | 73 | 1.18 | .27 | .90 | -.97 | .89 | -.78 | d | .48 | .40 | 75.8 | 68.1 | S18 |
| 17 | 57 | 73 | .52 | .30 | .89 | -.73 | .78 | -.96 | c | .44 | .34 | 78.8 | 76.2 | S17 |
| 15 | 58 | 73 | .43 | .31 | .88 | -.73 | .71 | -1.25 | b | .45 | .33 | 77.3 | 77.5 | S15 |
| 14 | 59 | 73 | .33 | .32 | .86 | -.79 | .71 | -1.16 | a | .45 | .32 | 81.8 | 78.9 | S14 |
| MEAN | 60.7 | 73.0 | .00 | .37 | 1.00 | .09 | .96 | .01 | | | | 82.4 | 82.1 | |
| P. SD | 6.1 | .0 | .76 | .09 | .09 | .64 | .28 | .88 | | | | 8.5 | 8.2 | |

Gambar 3. Konsistensi internal setiap butir pernyataan.

Analisis selanjutnya yang dilakukan pada kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik yang telah dikembangkan adalah analisis bias pernyataan. Bias yang dilakukan berfokus pada jenis kelamin responden saja. Hasil analisis Model Rasch menunjukkan bahwa nilai DIF tidak kurang dari 5% (0,05). Ini

menunjukkan bahwa tidak ada satu pun item pertanyaan dalam set instrumen tes diagnostik yang bias terhadap perbedaan jenis kelamin.

| Person CLASSES | SUMMARY DIF | | | BETWEEN-CLASS/GROUP | | | Item | |
|-------------------|-------------|------|-------|---------------------|-------|------|--------|------|
| | CHI-SQUARED | D.F. | PROB. | UNWTD | MNSQ | ZSTD | Number | Name |
| 10 | 6.8781 | 9 | .6498 | 1.2128 | .58 | | 1 | S1 |
| 10 | 7.8031 | 9 | .5541 | 1.4314 | .97 | | 2 | S2 |
| 10 | 7.8936 | 9 | .5448 | 1.6048 | 1.24 | | 3 | S3 |
| 10 | 3.2946 | 9 | .9514 | .7686 | -.38 | | 4 | S4 |
| 10 | 5.0514 | 9 | .8298 | 1.1800 | .52 | | 5 | S5 |
| 10 | 4.4031 | 9 | .8829 | 1.0104 | .18 | | 6 | S6 |
| 10 | 3.3000 | 9 | .9512 | .7288 | -.48 | | 7 | S7 |
| 10 | 4.3471 | 9 | .8871 | 1.0124 | .18 | | 8 | S8 |
| 10 | 4.0469 | 9 | .9083 | .8005 | -.30 | | 9 | S9 |
| 10 | 4.8800 | 9 | .8446 | .9214 | -.01 | | 10 | S10 |
| 10 | 5.2471 | 9 | .8122 | 1.1067 | .38 | | 11 | S11 |
| 10 | 5.6151 | 9 | .7777 | 1.0515 | .26 | | 12 | S12 |
| 10 | 2.5725 | 9 | .9789 | .4868 | -1.20 | | 13 | S13 |
| 10 | 4.3188 | 9 | .8892 | .8577 | -.16 | | 14 | S14 |
| 10 | 4.5699 | 9 | .8701 | .9108 | -.04 | | 15 | S15 |
| 10 | 3.9019 | 9 | .9177 | .8121 | -.27 | | 16 | S16 |
| 10 | 4.8014 | 9 | .8512 | .8606 | -.15 | | 17 | S17 |
| 10 | 2.9694 | 9 | .9655 | .4115 | -1.47 | | 18 | S18 |
| 10 | 6.7623 | 9 | .6618 | 1.1889 | .54 | | 19 | S19 |
| 10 | 6.5883 | 9 | .6799 | 1.0320 | .22 | | 20 | S20 |

Gambar 4. Nilai DIF setiap butir pernyataan.

Nilai DIF memeriksa kemungkinan item bias yang disebabkan oleh pola jawaban dari kelompok yang berbeda (Rahmani, 2018; Veas *et al.*, 2016; Xu *et al.*, 2020). Pertanyaan bias muncul karena ada dua kelompok berbeda dengan kemampuan yang sama, tetapi ada perbedaan yang signifikan ketika menjawab item pada instrumen (Alwi, 2017). Oleh karena itu, informasi apakah suatu butir mengalami bias atau tidak penting untuk dianalisis.

Analisis tambahan yang dilakukan pada penelitian ini adalah analisis responden. Hasil analisis Model Rasch dilanjutkan untuk menentukan konsistensi internal setiap responden dalam menjawab kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik yang telah dikembangkan. Responden dinyatakan memiliki konsistensi internal atau *fit* jika memenuhi salah satu dari tiga kondisi. Ketiga kondisi tersebut adalah nilai *Outfit Mean Square* antara 0,5 hingga 1,5; nilai *Outfit Z-Standard* antara -2,0 hingga 2,0; dan nilai *Point Measure Correlation* antara 0,4 hingga 0,85. Hasil analisis Model Rasch menunjukkan bahwa seluruh responden dinyatakan *fit* karena memenuhi setidaknya satu dari tiga persyaratan konsistensi internal.

| 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% | Person | |
|--------------|-------------|-------------|--------------|-------------|------------|-------|-------------|-------|----------------|---------|------------|------------|--------|-----|
| 6 | 103 | 25 | 52.03 | 2.65 | 1.45 | 2.19 | 1.43 | 2.04 | A | .14 | .30 | 24.0 | 38.4 | 06L |
| 10 | 99 | 25 | 49.26 | 2.63 | 1.34 | 1.80 | 1.44 | 2.12 | B | .24 | .30 | 16.0 | 38.1 | 10P |
| 11 | 98 | 25 | 48.56 | 2.63 | 1.12 | .70 | 1.11 | .61 | C | .13 | .29 | 40.0 | 38.1 | 11L |
| 4 | 95 | 25 | 46.45 | 2.68 | 1.06 | .37 | 1.08 | .43 | D | .06 | .28 | 40.0 | 38.9 | 04L |
| 3 | 106 | 25 | 54.19 | 2.71 | 1.06 | .35 | 1.02 | .18 | E | .19 | .30 | 44.0 | 40.3 | 03P |
| 5 | 97 | 25 | 47.87 | 2.64 | .91 | -.45 | .90 | -.48 | F | .35 | .29 | 48.0 | 39.2 | 05L |
| 8 | 97 | 25 | 47.87 | 2.64 | .90 | -.53 | .87 | -.62 | e | .54 | .29 | 40.0 | 39.2 | 08P |
| 9 | 93 | 25 | 44.98 | 2.73 | .90 | -.47 | .87 | -.55 | d | .50 | .28 | 44.0 | 39.7 | 09L |
| 1 | 98 | 25 | 48.56 | 2.63 | .81 | -1.05 | .80 | -1.06 | c | .40 | .29 | 48.0 | 38.1 | 01L |
| 7 | 102 | 25 | 51.33 | 2.64 | .80 | -1.11 | .78 | -1.19 | b | .20 | .30 | 48.0 | 37.5 | 07P |
| 2 | 97 | 25 | 47.87 | 2.64 | .69 | -1.83 | .68 | -1.83 | a | .50 | .29 | 56.0 | 39.2 | 02P |
| MEAN | 98.6 | 25.0 | 49.00 | 2.66 | 1.00 | .00 | 1.00 | -.03 | | | | 40.7 | 38.8 | |
| P. SD | 3.5 | .0 | 2.50 | .03 | .22 | 1.17 | .24 | 1.21 | | | | 10.9 | .8 | |

Gambar 5. Konsistensi internal setiap responden.

D. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah diuraikan, penelitian ini menghasilkan kesimpulan bahwa produk pendidikan berupa kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik dinyatakan layak dan dapat digunakan untuk implementasi pada masyarakat umum. Hal ini dibuktikan melalui hasil uji teoritis (CVR dan CVI) dan empiris (Rasch Model) yang menyatakan bahwa kuesioner validasi teoritik butir soal pilihan ganda pada instrumen tes diagnostik valid dari segi konstruk dan konstruk, serta reliabel.

Penelitian ini telah berhasil mengembangkan kuesioner validasi teoritik untuk butir soal pilihan ganda pada instrumen tes diagnostik. Adapun penelitian ini dapat dilanjutkan kembali untuk butir soal berbentuk uraian.

E. REFERENSI

Al-Taweel, D., & Awad, A. (2020). Development and validation of medication assessment tools to evaluate prescribing adherence to evidence-based guidelines for secondary prevention of coronary heart disease in post-acute coronary syndromes patients in Kuwait. *PLoS ONE*, 15(1), 1–16. <https://doi.org/10.1371/journal.pone.0241633>

Al Zoubi, F., Mayo, N., Rochette, A., & Thomas, A. (2018). Applying modern measurement approaches to constructs relevant to evidence-based



practice among Canadian physical and occupational therapists. *Implementation Science*, 13(1), 1–21. <https://doi.org/10.1186/s13012-018-0844-4>

Alizadeh-Siuki, H., Tehrani, H., Gholian-Aval, M., Ebrahimipour, H., Jafari, A., & Vahedian-Shahroodi, M. (2020). Psychometric properties of a questionnaire on brucellosis prevention behaviors based on the precede model among rural farmers and their family members. *Risk Management and Healthcare Policy*, 13(1), 539–548. <https://doi.org/10.2147/RMHP.S247356>

Alwi, I. (2017). Sensitivity of Mantel Haenszel Model and Rasch Model as Viewed from Sample Size. *Jurnal Evaluasi Pendidikan*, 2(1), 18–29. <https://doi.org/10.21009/jep.021.02>

Bodzin, A., Hammond, T., Fu, Q., & Farina, W. (2020). Development of Instruments to Assess Students' Spatial Learning Attitudes (SLA) and Interest in Science, Technology and Geospatial Technology (STEM-GEO). *International Journal of Educational Methodology*, 6(1), 67–81. <https://doi.org/10.12973/ijem.6.1.67>

Boone, W. J. (2016). Rasch Analysis for Instrument Development: Why, When, and How? *CBE Life Sciences Education*, 15(4), 1–7. <https://doi.org/10.1187/cbe.16-04-0148>

Chang, J., Li, Y., Song, H., Yong, L., Luo, L., Zhang, Z., & Song, N. (2020). Assessment of Validity of Children's Movement Skill Quotient (CMSQ) Based on the Physical Education Classroom Environment. *BioMed Research International*, 2020(1), 1–11. <https://doi.org/10.1155/2020/8938763>

Fulmer, G. W., Chu, H., Treagust, D. F., & Neumann, K. (2016). Is it harder to know or to reason? Analyzing two-tier science assessment items using the Rasch measurement model. *Asia-Pacific Science Education*, 1(2015), 1–16. <https://doi.org/10.1186/s41029-015-0005-x>

Islam, A. Y. M. A., Gu, X., Crook, C., & Spector, J. M. (2020). Assessment of ICT in Tertiary Education Applying Structural Equation Modeling and

Rasch Model. *SAGE Open*, 10(4), 1–17.
<https://doi.org/10.1177/2158244020975409>

Koçak, D. (2020). Investigation of rater tendencies and reliability in different assessment methods with many facet rasch model. *International Electronic Journal of Elementary Education*, 12(4), 349–358.
<https://doi.org/10.26822/iejee.2020459464>

Ling, M., Pang, V., & Salwana, B. (2014). Rasch Analysis: Psychometric Properties of Malaysian Secondary School Students' Leadership Scale. *Man In India*, 96(1–2), 1–8.

Lo, C., Liang, W., Hang, L., Wu, T., Chang, Y., & Chang, C. (2015). A Psychometric Assessment of the St. George's Respiratory Questionnaire in Patients with COPD using Rasch Model Analysis. *Health and Quality of Life Outcomes*, 13(131), 1–16. <https://doi.org/10.1186/s12955-015-0320-7>

Matore, M. E. E. M., Maat, S. M., Affandi, H. M., Mohamad, S., & Khairani, A. Z. (2018). Assessment of psychometric properties for Raven Advanced Progressive Matrices in measuring intellectual quotient (IQ) using Rasch model. *Asian Journal of Scientific Research*, 11(3), 393–400.
<https://doi.org/10.3923/ajsr.2018.393.400>

Md Yunos, J., Ibrahim Mukhtar, M., Alias, M., Lee, M. F., Tee, T. K., Rubani, S. N. K., Hamid, H., Yunus, F. A. N., Sulaiman, J., & Sumarwati, S. (2017). Validity of vocational pedagogy constructs using the rasch measurement model. *Journal of Technical Education and Training*, 9(2), 35–45.

Mohamad, M. M., Lisa, N., Sern, L. C., & Mohd, K. (2015). Measuring the Validity and Reliability of Research Instruments. *Procedia - Social and Behavioral Sciences*, 204(1), 164–171.
<https://doi.org/10.1016/j.sbspro.2015.08.129>

Mokshein, S. E., Ishak, H., & Ahmad, H. (2019). The use of rasch measurement model in English testing. *Cakrawala Pendidikan*, 38(1), 16–32. <https://doi.org/10.21831/cp.v38i1.22750>

- Nielsen, T., & Dammeyer, J. (2019). Measuring higher education students' perceived stress: An IRT-based construct validity study of the PSS-10. *Studies in Educational Evaluation*, 63(1), 17–25. <https://doi.org/10.1016/j.stueduc.2019.06.007>
- Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, 15(2), 1–14. <https://doi.org/10.1103/PhysRevPhysEducRes.15.020111>
- Rahmani, B. D. (2018). Differential Item Functional Analysis on Pedagogic and Content Knowledge (PCK) Questionnaire for Indonesian Teachers using RASCH Model. *Journal of Physics: Conference Series*, 948(1), 1–8.
- Semakula, D., Nsangi, A., Oxman, A. D., Sewankambo, N. K., Guttersrud, O., & Austvoll-Dahlgren, A. (2017). Measuring ability to assess claims about treatment effects in English and Luganda: evaluation of multiple-choice questions from the “Claim Evaluation Tools” database using Rasch modelling. *Informed Health Choices*, 1–17.
- Sjaastad, J. (2014). Enhancing measurement in science and mathematics education research through Rasch analysis: Rationale and properties. *NORDINA*, 10(2), 212–230.
- Suryani, Y. E. (2018). Aplikasi Rasch Model dalam Mengevaluasi Intelligenz Structure Test. *Psikohumaniora: Jurnal Penelitian Psikologi*, 3(1), 73–100.
- Susanti, D., Supriyati, Y., Maulana, S., & Hakim, A. R. (2021). Analysis of students' misconception and strategic thinking skill on parabolic motion material using two-tier multiple choice items instrument. *Journal of Physics: Conference Series*, 2019(1). <https://doi.org/10.1088/1742-6596/2019/1/012050>
- Veas, A., Castejón, J. L., Gilar, R., & Miñano, P. (2016). Validation of the Spanish adaptation of the School Attitude Assessment Survey-Revised using multidimensional Rasch analysis. *Anales de Psicología*, 33(1), 74–81. <https://doi.org/10.6018/analesps.33.1.235271>

- Waltz, C. F., Strickland, O. L., & Lenz, E. R. (2016). Measurement in Nursing and Health Research. In *Measurement in Nursing and Health Research*. <https://doi.org/10.1891/9780826170620>
- Wilantika, N., Khoiri, N., & Hidayat, S. (2018). Pengembangan penyusunan instrumen four-tier diagnostic test untuk mengungkap miskonsepsi materi sistem ekskresi di SMA Negeri 1 Mayong Jepara. *Phenomenon: Jurnal Pendidikan* ..., 8(2), 200–214. <https://103.19.37.186/index.php/Phenomenon/article/view/2699>
- Xiao, Y., Han, J., Koenig, K., Xiong, J., & Bao, L. (2018). Multilevel Rasch modeling of two-tier multiple choice test : A case study using Lawson ' s classroom test of scientific reasoning. *Physical Review Physics Education Research*, 14(2), 1–18. <https://doi.org/10.1103/PhysRevPhysEducRes.14.020104>
- Xu, R. H., Wong, E. L. yi, Lu, S. Y. jun, Zhou, L. M., Chang, J. H., & Wang, D. (2020). Validation of the Toronto Empathy Questionnaire (TEQ) Among Medical Students in China: Analyses Using Three Psychometric Methods. *Frontiers in Psychology*, 11(1), 1–11. <https://doi.org/10.3389/fpsyg.2020.00810>
- Yasin, R. M., Yunus, F. A. N., Rus, R. C., Ahmad, A., & Rahim, M. B. (2015). Validity and Reliability Learning Transfer Item Using Rasch Measurement Model. *Procedia - Social and Behavioral Sciences*, 204(1), 212–217. <https://doi.org/10.1016/j.sbspro.2015.08.143>
- Yasin, S. N. T. M., Yunus, M. F. M., & Ismail, I. (2018). The use of rasch measurement model for the validity and reliability. *Journal of Counseling and Educational Technology*, 1(2), 22–27. <https://doi.org/10.32698/0111>
- Zamri, A., & Nordin. (2015). Modeling a Multiple Choice Mathematics Test with the Rasch Model. *Indian Journal of Science and Technology*, 8(12), 1–6. <https://doi.org/10.17485/ijst/2015/v8i12/70650>