



Analisis kualitas butir soal ujian akhir semester mata pelajaran biologi SMA menggunakan Rasch Model

Opik Prasetyo ^{1,*}

^a SMA Pradita Dirgantara, Kabupaten Boyolali, Jawa Tengah, 57375, Indonesia.

¹ opik.prasetyo@praditadirgantara.sch.id*

* Corresponding author

INFORMASI ARTIKEL

Lini Masa Artikel		Keywords
Draft diterima	: 2023-10-12	<i>Analysis of the quality of test item;</i>
Revisi diterima	: 2023-12-02	<i>Biology learning;</i>
Diterbitkan	: 2024-04-15	<i>Rasch model;</i>
ABSTRAK		ABSTRACT
Penelitian ini bertujuan untuk mengetahui kualitas butir soal ujian akhir semester mata pelajaran Biologi SMA menggunakan RASCH model. Soal UAS yang dianalisis berjumlah 30 soal pilihan ganda pada materi ekologi dan sel. Responden merupakan 50 siswa kelas XI SMA tahun akademik 2021/2022. Teknik analisis data dalam penelitian ini adalah <i>Item Responses Theory (IRT)</i>. Hasil analisis menunjukkan bahwa dari 30 butir soal, sejumlah 17 item termasuk kategori valid, sedangkan 13 item lainnya invalid. Nilai Alpha Cronbach berdasarkan hasil analisis adalah $0,69 < 0,7$; menunjukkan bahwa beberapa soal yang invalid tersebut tidak reliabel dan memerlukan perbaikan, sehingga keajegan atau konsistensi hasil pengukuran menggunakan item soal menjadi lebih dapat dipercaya. Hasil analisis juga menunjukkan bahwa mayoritas siswa memiliki kesesuaian dengan RASCH model yang dikembangkan, kecuali siswa nomor 04L, 18P, 07P, 44P, 39L, 43P dan 25P.		This research uses the Rasch model to determine the quality of Biology final exam questions in senior high school. The final exam questions consist of 30 multiple-choice questions on ecology and cells. The answers are collected from 50 students in grade XI in SMA Academic Year 2021/2022. This research's data analysis technique is the Item Response Theory (IRT). The result shows that from 30 questions, 17 are in the valid category, while the other 13 are invalid. The value of Cronbach's alpha is $0.69 < 0.7$, which shows that some invalid questions are unreliable and require improvement. Hence, revising the items to enhance their clarity, relevance, and alignment with the intended constructs improves the consistency of the measurement results. The analysis also shows that most students conform to the developed Rasch model, except for student numbers 04L, 18P, 07P, 44P, 39L, 43P, and 25P.

Cara Sitasi Artikel Ini (APA Style):

Prasetyo, O. (2024). Analisis kualitas butir soal ujian akhir semester mata pelajaran biologi SMA menggunakan Rasch Model. *Bio-Pedagogi*. 13(1), 15-27. <https://dx.doi.org/10.20961/bio-pedagogi.v13i1.79390>.

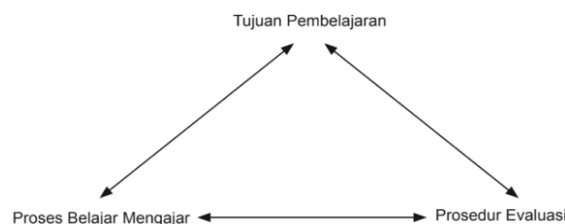
Artikel ini dapat diakses secara bebas dengan lisensi [C C-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/).



PENDAHULUAN

Evaluasi merupakan bagian integral untuk menentukan berhasil tidaknya kegiatan yang telah dilaksanakan. Evaluasi adalah proses penilaian atau pengukuran terhadap suatu kegiatan, proyek, program, atau sistem guna mengevaluasi efektivitas, efisiensi, dan dampaknya. Evaluasi bertujuan untuk mendapatkan informasi yang objektif dan akurat tentang kinerja suatu hal, serta memberikan dasar untuk pengambilan keputusan dan perbaikan di masa depan. Evaluasi merupakan alat penting dalam pengambilan keputusan yang didasarkan pada bukti-bukti objektif. Dengan melakukan evaluasi secara teratur, organisasi atau individu dapat meningkatkan kinerja mereka, mengidentifikasi masalah, dan mengoptimalkan penggunaan sumber daya.

Dalam konteks umum, evaluasi dapat dilakukan dalam berbagai bidang, seperti pendidikan, bisnis, pemerintahan, lingkungan, dan kesehatan. Evaluasi dapat melibatkan pengumpulan data, analisis data, pengukuran kinerja, penilaian efisiensi, dan penilaian dampak. Kegiatan evaluasi yang dilakukan oleh guru memiliki berbagai macam tujuan dan manfaat ([Wibisono, 2018](#)). Proses evaluasi bertujuan untuk mengumpulkan berbagai bahan ataupun informasi yang dapat dijadikan sebagai *evidence* untuk mengetahui tingkat perkembangan peserta didik setelah mengikuti aktivitas belajar mengajar selama periode waktu tertentu, serta untuk mengukur keefektifan metode pengajaran yang diterapkan oleh guru ([Sudijono, 2008](#)). Evaluasi pembelajaran juga bermanfaat untuk mengetahui efektivitas dan efisiensi proses pembelajaran baik dari segi penyusunan tujuan pembelajaran, metode pengajaran, materi, media dan sumber belajar, serta sistem asesmen yang diterapkan dalam pembelajaran tersebut ([Febriyanti et al., 2022](#); [Magdalena & Apriliyani, 2023](#)). Kegiatan evaluasi di dalam dunia pendidikan harus dilakukan dengan sistematis, terencana dan berkesinambungan ([Olson & Bakken, 2017](#)). Kegiatan ini juga tidak lepas dari tujuan pembelajaran yang telah dirumuskan oleh guru. Oleh karena itu, kegiatan evaluasi sangat berkaitan erat dengan pelaksanaan proses pembelajaran dan pencapaian tujuan pembelajaran yang telah dirumuskan (**Gambar 1**).



Gambar 1. Proses Belajar Mengajar

Kegiatan evaluasi terdiri dari proses pengukuran dan penilaian oleh pendidik terhadap peserta didik yang telah mengikuti rangkaian proses pembelajaran yang diselenggarakan ([Sudijono, 2008](#)). Kedua proses ini, yakni pengukuran dan penilaian memiliki kedudukan yang penting dan tak terpisahkan dari suatu kegiatan evaluasi pembelajaran.

Pengukuran (*measurement*) merupakan proses pemberian angka bagi peserta didik yang diharapkan mampu mencerminkan kemampuannya dalam suatu mata pelajaran ([Weisburd & Britt, 2014](#)). Pengukuran adalah proses atau metode untuk memperoleh informasi kuantitatif atau numerik mengenai suatu objek, peristiwa, atau fenomena. Pengukuran dilakukan dengan menggunakan alat pengukur atau instrumen tertentu untuk mengumpulkan data yang dapat dinyatakan dalam bentuk angka atau satuan yang dapat diukur. Dalam proses pengukuran, dibutuhkan alat ukur yang valid dan reliabel guna memberikan informasi yang akurat untuk menggambarkan kemampuan peserta didik dalam bidang pelajaran tertentu ([Wibowo dan Cholifah, 2018](#)). Kegiatan pengukuran memiliki beberapa karakteristik antara lain, terdapat atribut yang akan diukur, hasil pengukuran dinyatakan secara kuantitatif atau berupa angka, dan hasil bersifat deskripsi, yakni hanya memberikan angka yang mana tidak diterjemahkan lebih lanjut ([Guttsayt, 2020](#); [Pospíšil et al., 2022](#)).

Selain proses pengukuran, di dalam melaksanakan kegiatan evaluasi pembelajaran juga dilakukan proses penilaian (*assessment*). Melalui kegiatan penilaian inilah data hasil pengukuran akan diinterpretasikan. Penilaian adalah serangkaian proses untuk mendapatkan dan memberikan informasi yang dapat digunakan sebagai rujukan alternatif untuk membuat keputusan ([Ademi, 2017](#)). Proses penilaian ini meliputi pengumpulan bukti-bukti yang menggambarkan pencapaian hasil belajar peserta didik ([Reotutar et al., 2020](#)). Penilaian adalah proses untuk mengevaluasi atau menilai kualitas, nilai, atau keberhasilan suatu objek, situasi, atau individu berdasarkan kriteria atau standar yang telah ditetapkan.

Penilaian dapat dilakukan secara subjektif atau objektif, tergantung pada metode dan kriteria yang digunakan. Tujuan dari penilaian adalah untuk memberikan umpan balik, mengidentifikasi kekuatan dan kelemahan, mengukur pencapaian, dan membuat keputusan berdasarkan hasil penilaian tersebut. Berbagai penelitian menunjukkan bahwa penggunaan metode pengukuran dan penilaian berorientasi proses memungkinkan pemantauan kemajuan peserta didik dan dalam hal ini, guru dapat memenuhi kebutuhan mereka peserta didik dalam belajar ([Vaiz, Osman, et al, 2021](#)).

Terdapat berbagai bentuk pengukuran dan penilaian yang dilakukan oleh pendidik untuk mengevaluasi proses pengajaran yang dilakukan, antara lain *self-assessment*, portofolio, penilaian kerja, *peer review*, *concept map*, *crossword*, *project*, dan berbagai jenis tes lainnya ([Ziborova et al., 2020](#)). Bentuk pengukuran dapat bervariasi tergantung pada jenis variabel yang diukur dan tujuan dari pengukuran tersebut. Berikut ini beberapa bentuk umum pengukuran yang digunakan dalam penelitian dan evaluasi menurut [Dombi & Jónás \(2021\)](#) :

1. Skala Likert: Skala Likert adalah bentuk pengukuran yang digunakan untuk mengukur tingkat persetujuan, pendapat, atau sikap seseorang terhadap pernyataan tertentu. Responden diminta untuk menunjukkan tingkat persetujuan mereka dengan menyatakan sejauh mana mereka setuju atau tidak setuju terhadap pernyataan yang diberikan, biasanya dalam skala 5 atau 7 poin.
2. Skala Interval: Skala interval adalah bentuk pengukuran yang menggunakan angka-angka yang memenuhi syarat matematika untuk mengukur properti seperti suhu, waktu, atau jarak. Skala ini memiliki titik nol yang absolut dan memungkinkan perhitungan perbedaan antara nilai-nilai.
3. Skala Ordinal: Skala ordinal adalah bentuk pengukuran yang digunakan untuk mengurutkan variabel dalam urutan tertentu, tanpa memberikan informasi tentang perbedaan atau jarak antara nilai-nilai. Contoh umum adalah peringkat, seperti peringkat kejuaraan atau peringkat preferensi.
4. Skala Nominal: Skala nominal adalah bentuk pengukuran yang digunakan untuk mengelompokkan atau mengategorikan variabel dalam kategori-kategori yang saling eksklusif. Contoh umum adalah jenis kelamin, agama, atau status perkawinan.
5. Pengukuran Verbal: Pengukuran verbal melibatkan deskripsi atau narasi kualitatif tentang suatu fenomena. Ini digunakan ketika aspek yang diukur sulit diungkapkan dalam angka atau kategori.
6. Tes atau Soal: Tes atau soal adalah bentuk pengukuran yang digunakan untuk mengukur pemahaman, pengetahuan, atau keterampilan melalui pertanyaan atau pernyataan yang dijawab oleh peserta dengan benar atau salah.
7. Pengukuran Fisik: Pengukuran fisik melibatkan penggunaan instrumen atau peralatan khusus untuk mengukur variabel fisik seperti berat, tinggi, kecepatan, atau tekanan.

Pilihan bentuk pengukuran tergantung pada sifat variabel yang diukur dan tujuan pengukuran. Penting untuk memilih bentuk pengukuran yang paling sesuai dan relevan untuk mencapai tujuan penelitian atau evaluasi ([Wuang et al., 2009](#)). Proses penilaian merupakan sesuatu yang sangat penting dalam menentukan seberapa banyak siswa telah belajar di akhir periode proses pendidikan, juga untuk sekolah dan siswa. Proses ini memberikan umpan balik penting bagi siswa dan pendidik ([Chaudhary & Dey, 2013](#)). Proses penilaian memiliki beberapa alasan mengapa penting dalam berbagai konteks. Berikut ini adalah beberapa alasan mengapa proses penilaian dianggap penting menurut [Fachri \(2018\)](#) :

1. Pengukuran Kemajuan: Penilaian memungkinkan untuk mengukur kemajuan dan pencapaian seseorang, kelompok, atau program dalam mencapai tujuan yang ditetapkan. Dengan melakukan penilaian secara teratur, kita dapat melihat perkembangan dari waktu ke waktu dan menilai sejauh mana kemajuan telah dicapai.
2. Pengambilan Keputusan: Penilaian memberikan informasi untuk pengambilan keputusan yang objektif. Dengan data dan informasi yang akurat tentang kinerja atau kondisi yang sedang dievaluasi, keputusan yang diambil akan lebih berdasarkan pada fakta dan bukti.
3. Perbaikan dan Pengembangan: Proses penilaian memungkinkan identifikasi kelemahan, kekuatan, dan area pengembangan yang perlu diperhatikan. Informasi yang diperoleh dari penilaian dapat digunakan untuk melakukan perbaikan dan pengembangan di masa depan. Ini dapat berlaku untuk individu, organisasi, atau program.
4. Akuntabilitas dan Transparansi: Penilaian membantu menciptakan akuntabilitas dan transparansi dalam berbagai konteks, seperti dalam pendidikan, pemerintahan, dan bisnis. Dengan melakukan penilaian, pihak yang bertanggung jawab akan terpapar terhadap hasil kinerja mereka dan dapat bertanggung jawab atas hasil yang dicapai.
5. Peningkatan Kualitas: Proses penilaian yang baik membantu meningkatkan kualitas baik secara individu maupun kolektif. Dengan mengevaluasi kinerja, praktik, atau program, kita dapat mengidentifikasi peluang perbaikan dan menerapkan langkah-langkah yang diperlukan untuk meningkatkan kualitas.
6. Pengakuan dan Penghargaan: Penilaian dapat digunakan sebagai dasar untuk pengakuan dan penghargaan terhadap pencapaian yang luar biasa. Hasil penilaian yang positif dapat memberikan penghargaan dan motivasi bagi individu atau kelompok yang berkinerja baik.

Dalam keseluruhan pendidikan, proses penilaian penting karena memberikan informasi yang akurat dan objektif, membantu pengambilan keputusan, meningkatkan kualitas, menciptakan akuntabilitas, dan memberikan dasar untuk pengembangan di masa depan. Tes merupakan sekumpulan pertanyaan yang diajukan dan harus dijawab guna mengukur bagaimana kompetensi individu atau mengungkapkan beberapa aspek dari responden yang diuji ([Rapono et al., 2019](#)). Tes merupakan suatu alat atau cara yang diterapkan guna mengetahui sesuatu berdasarkan aturan yang telah disusun ([Akbaev & Babichev, 2023](#)). Kualitas *instrumen* tes yang baik sebagai alat ukur dalam pembelajaran akan menentukan seberapa baik dan valid data hasil penilaian untuk evaluasi.

Salah satu bentuk tes di dalam dunia Pendidikan adalah Ujian Akhir Semester (UAS). UAS dilakukan pada akhir periode atau semester akademik untuk menilai pemahaman dan pencapaian siswa dalam suatu mata pelajaran atau program studi ([Oermann, 2018](#)). Hasil UAS digunakan untuk menggambarkan pencapaian individu dalam mengikuti proses pendidikan tertentu, dan menentukan standar kualitas desain pengajaran yang diselenggarakan oleh guru. Oleh karena itu, instrumen tes yang digunakan harus memiliki bentuk pertanyaan yang berkualitas dan mampu mengukur semua kompetensi yang telah dicapai oleh peserta didik secara riil ([Efendi et al., 2019](#)).

Dalam menyusun soal tes atau ujian sebagai suatu *instrumen* penilaian, beberapa parameter harus diperhatikan oleh guru. Validitas, reliabilitas soal, indeks kesukaran (*difficulty index*), indeks daya beda, keefektifan distraktor merupakan parameter yang harus dipenuhi dalam menyusun dan menghasilkan instrumen tes yang berkualitas baik ([Maba et al., 2018](#)). Validitas adalah pengukuran harus valid, artinya instrumen pengukuran harus secara akurat mengukur apa yang sebenarnya ingin diukur. Validitas pengukuran dapat dikaitkan dengan sejauh mana instrumen pengukuran tersebut dapat mengukur konsep atau variabel yang diinginkan. Reliabilitas memberikan hasil yang konsisten dan dapat diandalkan dalam pengukuran berulang. Reliabilitas pengukuran mengacu pada tingkat ketepatan dan konsistensi instrumen pengukuran ([Yusup, 2018](#)).

Indeks kesukaran (*difficulty index*) adalah ukuran yang digunakan untuk mengevaluasi tingkat kesulitan atau tingkat kesukaran suatu soal dalam tes. Indeks kesukaran biasanya dinyatakan dalam angka antara 0 hingga 1, di mana nilai yang lebih tinggi menunjukkan tingkat kesulitan yang lebih tinggi ([Kaushik & Jayaram, 2016](#)). Indeks daya beda (*discrimination index*) adalah ukuran yang

digunakan untuk mengevaluasi sejauh mana suatu soal dalam tes dapat membedakan antara individu yang memiliki kemampuan tinggi dan rendah. Indeks daya beda biasanya dinyatakan dalam angka antara -1 hingga +1, di mana nilai yang lebih tinggi menunjukkan tingkat beda yang lebih baik ([Metsämuuronen, 2020](#)). Distraktor adalah pilihan jawaban yang salah atau tidak tepat dalam suatu soal pilihan ganda. Keefektifan distraktor merujuk pada sejauh mana distraktor tersebut dapat membedakan antara peserta yang memiliki pemahaman yang benar dengan peserta yang tidak memiliki pemahaman yang benar ([Rahma et al., 2017](#)).

Guru sangat memerlukan hal tersebut dalam mengetahui representasi nyata akan keberhasilan proses belajar mengajar yang diselenggarakan serta kondisi dan perkembangan masing – masing individu di dalam kelas. Penyusunan soal ataupun instrumen yang baik, dengan tingkat validitas dan reliabilitas yang tinggi adalah tantangan tersendiri bagi seorang guru. Berbagai hasil penelitian menunjukkan beberapa fakta bahwa terdapat soal – soal tertentu yang disusun oleh guru belum memenuhi kriteria soal yang baik sesuai parameter yang ditetapkan. Hal ini menunjukkan bahwa masih dibutuhkan adanya upaya untuk meningkatkan pemahaman guru akan karakteristik instrumen yang baik dan berkualitas.

Seperangkat instrumen soal yang baik memiliki berbagai bentuk karakteristik, antara lain tingkat kesukaran soal tersusun atas soal tipe mudah, sedang dan sulit harus tersebar dengan baik sesuai kebutuhan. Selain itu, abilitas belajar responden yang tinggi, sedang, dan rendah juga dapat dibedakan menggunakan instrumen tes yang berkualitas. Di sisi lain, validitas dan reliabilitas soal juga tidak dapat diabaikan guna mengetahui apakah item butir soal tersebut mampu mengukur kemampuan peserta didik secara tepat atau tidak. Untuk mengetahui karakteristik suatu butir soal, terdapat dua Teknik yang dapat dipakai, yaitu Rasch model ([Alavi et al., 2020](#)).

METODE

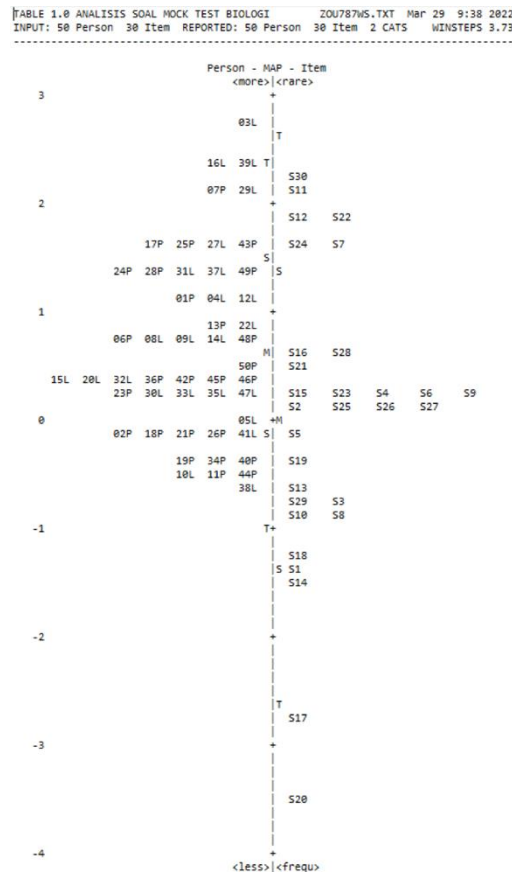
Penelitian ini merupakan penelitian deskriptif kuantitatif untuk memeriksa kualitas *instrument* soal ujian akhir semester mata pelajaran Biologi pada materi ekologi dan sel. Tahapan penelitian yang dilakukan yaitu: 1) penyusunan 30 butir soal pilihan ganda, 2) melaksanakan pengujian *instrument* soal kepada 50 siswa sebagai responden, 3) menganalisis hasil pengujian *instrument* soal menggunakan program *winstep*.

Sampel dalam penelitian ini yaitu responden merupakan 50 siswa kelas XI SMA salah satu sekolah di Solo Raya tahun akademik 2021/2022. Dalam penelitian ini peneliti memilih sampel yaitu siswa kelas XI yang telah menuntaskan materi ekologi dan sel. Teknik pengumpulan data dilakukan melalui analisis dokumen penilaian yang telah digunakan. Teknik analisis data dilakukan menggunakan pemodelan RASCH dengan menggunakan program *winstep*. Uji RASCH juga dikenal sebagai Analisis RASCH, adalah metode statistik yang digunakan untuk mengukur tingkat kemampuan atau karakteristik individu berdasarkan respons yang diberikan terhadap serangkaian item atau pertanyaan dalam suatu tes atau kuesioner. Tujuan dari Uji RASCH adalah untuk mengukur tingkat kemampuan individu secara objektif dan mengukur tingkat kesulitan atau daya pembeda item secara objektif. Analisis data berkaitan dengan validitas butir soal, reliabilitas soal, daya pembeda, dan tingkat kesukaran soal.

HASIL DAN PEMBAHASAN

Hasil penelitian menggunakan uji RASCH digunakan untuk mengetahui kesesuaian antara kemampuan siswa dengan model soal yang dikembangkan. Analisis model RASCH dalam penelitian ini digunakan untuk menganalisis soal UAS Biologi berjumlah 30 item soal dengan jumlah sampel sebanyak 50 siswa. Hasil analisis menunjukkan kualitas butir soal yang ditinjau dari aspek tingkat kesukaran, validitas, dan reliabilitas.

Tingkat Kesukaran Soal



Gambar 2. Peta konstruk dengan *winstep*

Hasil analisis berdasarkan interaksi antara item soal dan person pada **Gambar 2** menunjukkan soal dengan nilai logit tertinggi (2,26) yaitu nomor 30 adalah kategori soal paling sukar (gambar 3). Soal dengan nilai logit tertinggi adalah soal dengan tingkat kesulitan yang paling tinggi. Siswa yang berpotensi menjawab benar berdasarkan hasil analisis adalah siswa yang berinisial 03L, 16L, dan 39L. Menurut ([Hingorjo & Jaleel, 2012](#)), jika jumlah siswa yang dapat menjawab benar dibawah 30% dari total populasi, maka soal tersebut tergolong dalam kategori sulit. Berdasarkan hasil analisis, terdapat 6 soal yang tergolong dalam kategori sulit pada instrumen yang dikembangkan. Soal-soal tersebut adalah soal dengan inisial S24, S7, S12, S22, S11, and S30.

30. The table shows the monthly CO₂ concentrations in mg L⁻¹ taken at two monitoring stations.

Month Station	Jul 2011	Aug 2011	Sept 2011	Oct 2011	Nov 2011	Dec 2011	Jan 2012	Feb 2012	Mar 2012	Apr 2012	May 2012	Jun 2012
Cape Grim, Australia	388	389	389	389	389	389	389	389	389	389	389	390
Mauna Loa, Hawaii, USA	392	390	389	389	390	392	393	394	394	396	397	396

[Source: © International Baccalaureate Organization 2015]

What is directly indicated by the data?

- CO₂ concentration in the atmosphere varies from place to place.
- Cape Grim is less affected by global warming than Mauna Loa.
- CO₂ creates a greenhouse effect at both locations.
- The standard deviation for Cape Grim is higher than standard deviation for Mauna Loa.

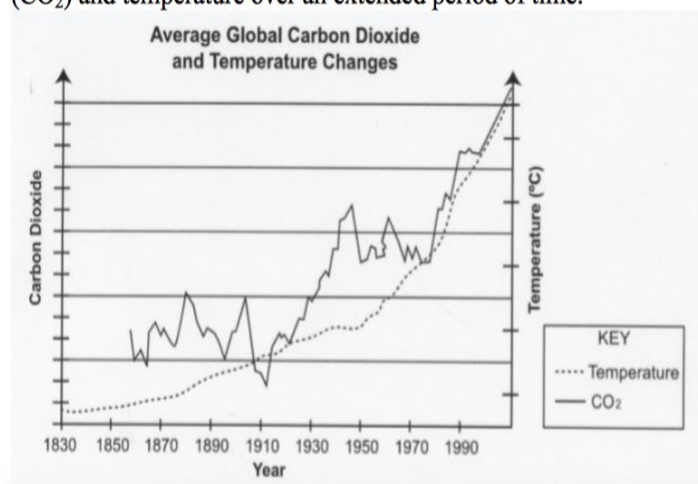
Gambar 3. Tipe Soal Kategori Sulit

Berdasarkan hasil analisis pada **Gambar 3**, dapat diidentifikasi soal dengan nilai logit paling rendah (-3,50) yaitu soal no 20 (S20). Hal ini menunjukkan bahwa soal tersebut adalah kategori soal

yang paling mudah (**Gambar 4**). Hasil analisis juga menunjukkan bahwa semua siswa berpotensi dapat menjawab benar pada soal tersebut. Menurut ([Hingorjo & Jaleel, 2012](#)), jika jumlah siswa yang dapat menjawab benar pada soal tersebut lebih dari 70%, maka soal tersebut termasuk dalam kategori mudah. Terdapat 21 soal yang termasuk dalam kategori mudah pada instrumen yang dikembangkan yaitu S20, S17, S14, S1, S18, S10, S8, S29, S3, S13, S9, S5, S2, S25, S26, S27, S15, S23, S24, S6 and S19.

Read the following article for questions no 19-20.

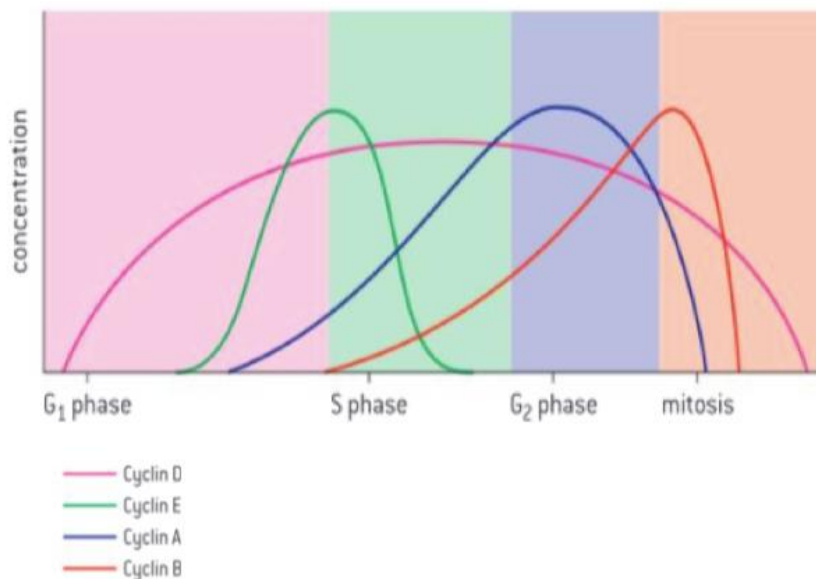
The burning of fossil fuels to heat homes, power factories, and run automobiles is largely responsible for increasing carbon dioxide (CO₂) emissions. Many scientists hypothesize that the increase in these greenhouse gases contributes directly to climate change (i.e., global warming). The graph below shows average global changes in carbon dioxide (CO₂) and temperature over an extended period of time.



Gambar 4. Tipe soal Kategori Mudah

Selanjutnya, hasil analisis pada soal dengan kategori medium. Tipe soal yang termasuk dalam kategori medium adalah soal yang dapat dijawab dengan benar oleh siswa sejumlah 30 - 70% dari total populasi ([Hingorjo & Jaleel, 2012](#)). Berdasarkan peta konstruk yang terdapat pada gambar 2, terdapat 3 soal yang termasuk dalam kategori medium. Soal-soal yang termasuk dalam kategori medium yaitu S16, S28 and S21 (**Gambar 5**). Soal-soal ini bernilai logit antara 0 sampai dengan 1.

21. Cyclin is a group of protein used to ensure many important task in cell cycle. The graph bellow shows how the levels of four main types of cyclin rise and fall in human cell. What will more likely to happen if cyclin D is not produced by the cell or maintained to have very low concentration?



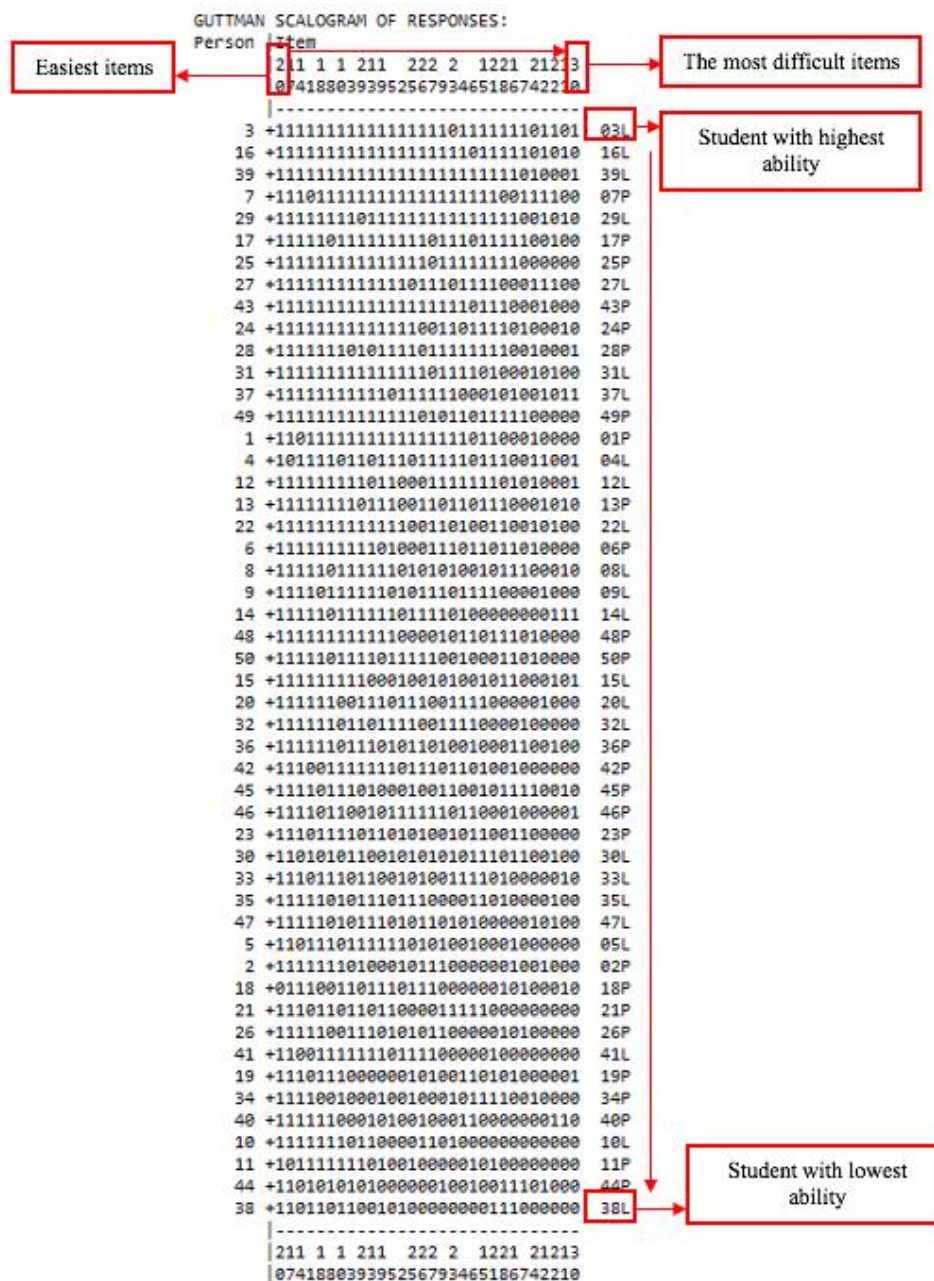
- Cell can not activate DNA replication inside the nucleus in S phase.
- Mitotic spindle will be formed but in very low level.
- Cell can not be triggerred from S to G2 phase.
- Cell will be at rest state

Gambar 5. Tipe Soal Kategori Medium

Selain kategori sulit, mudah, dan medium, terdapat satu kategori lainnya yaitu soal *item free person*. Soal yang tergolong dalam *item free person* adalah tipe soal yang tidak memiliki objektivitas yang spesifik dalam membedakan kemampuan siswa (Stemler & Naples, 2021). Artinya semua siswa dalam satu populasi dalam menjawab soal tersebut dengan benar atau semuanya tidak dapat menjawab sama sekali. Pada instrumen yang dikembangkan, terdapat beberapa soal yang termasuk dalam kategori soal *item free person*. Soal yang termasuk kategori *item free person* adalah S20 (Figure 4), S17, S14, S1, S18, S10, S8, S29, S3. Soal-soal tersebut memerlukan revisi karena termasuk dalam kategori soal mudah dan soal *item free person*.

Pola Jawaban Siswa

Informasi mengenai pola jawaban siswa pada instrumen yang dikerjakan dapat terlihat pada scalogram (Gutman matrix) pada **Gambar 6**. Data pada scalogram menunjukkan urutan kemampuan siswa yang sistematis dari yang abilitasnya tinggi (terdapat pada sisi atas) hingga abilitas rendah (terdapat pada sisi bawah). Selain itu, juga menggambarkan pola distribusi soal dari yang paling mudah (terdapat pada sisi kiri) hingga yang paling sulit (terdapat pada sisi paling kanan). Pola ini memudahkan peneliti atau guru dalam mengidentifikasi siswa yang menguasai sebagian besar kompetensi dan siswa yang mengalami kesulitan dalam menjawab soal. Selain itu, posisi jawaban yang tidak konsisten terhadap pola umum dapat menunjukkan adanya dugaan guessing atau miskonsepsi pada butir tertentu, sehingga dapat dijadikan dasar untuk mengevaluasi butir soal maupun strategi pembelajaran yang digunakan.



Gambar 6. Pola Jawaban Siswa pada Scalogram

Data pada **Gambar 6** menunjukkan pola distribusi kemampuan siswa dari yang tertinggi (03L) hingga yang terendah (38L). Siswa dengan kemampuan paling tinggi dapat menjawab benar 27 soal dari 30 soal yang tersedia. Sedangkan siswa dengan kemampuan paling rendah mampu menjawab dengan benar sebanyak 11 soal dari 30 soal yang tersedia. Jawaban yang benar direpresentasikan dengan angka satu, sedangkan jawaban yang salah direpresentasikan dengan angka nol.

Kesesuaian Siswa dengan Pemodelan Rasch

Kesesuaian antara abilitas siswa dengan model Rasch yang dikembangkan terdapat pada **Gambar 7**. Parameter yang digunakan yaitu *Outfit mean square* (OUTFIT MNSQ), *Outfit Z-standard* (OUTFIT ZSTD), dan *Point measure correlation* (PTMEASURE). Nilai *Outfit mean square* (OUTFIT MNSQ), *Outfit Z-standard* (OUTFIT ZSTD), dan *Point measure correlation* (PTMEASURE) pada instrumen yang dikembangkan antara terdapat pada **Tabel 1**.

Tabel 1. Item Measure

INPUT: 50 Person 30 Item REPORTED: 50 Person 30 Item 2 CATS WINSTEPS 3.73

Person: REAL SEP.: 1.49 REL.: .69 ... Item: REAL SEP.: 3.08 REL.: .90

Item STATISTICS: MEASURE ORDER

ENTRY	TOTAL	TOTAL	MEASURE	MODEL			INFIT			OUTFIT		PT-MEASURE		EXACT	MATCH	Item
NUMBER	SCORE	COUNT		S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%				
30	10	50	2.26	.38	1.06	.3	1.08	.3	.27	.33	82.0	80.7	S30			
11	11	50	2.12	.36	1.18	.9	1.28	1.0	.13	.34	76.0	79.1	S11			
12	13	50	1.87	.35	.93	-.3	.99	.0	.41	.35	80.0	76.2	S12			
22	13	50	1.87	.35	1.16	.9	1.19	.8	.17	.35	72.0	76.2	S22			
7	15	50	1.65	.33	1.18	1.2	1.28	1.3	.15	.36	72.0	73.3	S7			
24	15	50	1.65	.33	1.10	.7	1.05	.3	.27	.36	68.0	73.3	S24			
16	25	50	.66	.30	1.20	1.9	1.20	1.5	.15	.37	50.0	65.8	S16			
28	26	50	.57	.30	1.08	.8	1.08	.7	.28	.36	64.0	65.6	S28			
21	27	50	.48	.30	.84	-1.6	.81	-1.5	.54	.36	80.0	65.5	S21			
4	29	50	.29	.31	.84	-1.7	.79	-1.5	.54	.35	78.0	65.8	S4			
6	29	50	.29	.31	1.29	2.6	1.34	2.1	.03	.35	46.0	65.8	S6			
15	29	50	.29	.31	.86	-1.4	.81	-1.3	.52	.35	70.0	65.8	S15			
9	30	50	.20	.31	.85	-1.5	.79	-1.4	.53	.35	66.0	66.2	S9			
23	30	50	.20	.31	.96	-.3	1.09	.6	.37	.35	74.0	66.2	S23			
2	31	50	.10	.31	.91	-.9	.83	-1.0	.47	.34	72.0	67.0	S2			
25	31	50	.10	.31	1.04	.4	.99	.0	.31	.34	64.0	67.0	S25			
26	31	50	.10	.31	1.06	.6	1.01	.1	.29	.34	64.0	67.0	S26			
27	31	50	.10	.31	1.01	.2	1.00	.0	.33	.34	64.0	67.0	S27			
5	33	50	-.09	.32	.83	-1.5	.75	-1.4	.54	.33	74.0	68.7	S5			
19	36	50	-.41	.33	.91	-.6	.77	-1.0	.44	.31	72.0	72.5	S19			
13	38	50	-.64	.35	.97	-.1	.94	-.1	.32	.29	78.0	76.0	S13			
3	39	50	-.76	.36	.94	-.2	.82	-.5	.37	.28	78.0	78.0	S3			
29	39	50	-.76	.36	.94	-.3	1.02	.2	.34	.28	78.0	78.0	S29			
8	40	50	-.89	.37	1.03	.2	1.00	.1	.24	.27	80.0	80.0	S8			
10	40	50	-.89	.37	.98	.0	.83	-.4	.32	.27	80.0	80.0	S10			
18	42	50	-1.19	.40	1.00	.1	.81	-.4	.28	.25	84.0	84.0	S18			
1	43	50	-1.36	.42	1.03	.2	1.26	.7	.17	.23	86.0	86.0	S1			
14	44	50	-1.54	.45	.94	-.1	.80	-.3	.30	.22	88.0	88.0	S14			
17	48	50	-2.77	.73	1.01	.2	1.23	.5	.09	.13	96.0	96.0	S17			
20	49	50	-3.50	1.02	1.00	.3	.59	.1	.14	.09	98.0	98.0	S20			
MEAN	30.6	50.0	.00	.37	1.00	.0	.98	.0			74.5	74.6				
S.D.	10.8	.0	1.32	.14	.12	1.0	.19	.9			11.0	9.0				

Tabel 1 menunjukkan tipe soal yang tergolong dalam kategori sulit untuk kemampuan rata-rata siswa. Soal-soal tersebut memiliki nilai logit di atas 0,00, yaitu antara 1,65 hingga 2,26. Sedangkan soal-soal dengan nilai logit antara 0,48 hingga 0,66 tergolong dalam kategori medium. Sementara soal-soal yang memiliki nilai logit antara -3,50 hingga 0,29 termasuk dalam kategori mudah.

Kesesuaian antara item yang dikembangkan dengan kemampuan siswa dilihat dari tiga parameter yaitu *Outfit mean square* (OUTFIT MNSQ), *Outfit Z-standard* (OUTFIT ZSTD), dan *Point measure correlation* (PTMEASURE). Kriteria pertama adalah *Outfit mean square* (OUTFIT MNSQ). Hasil analisis terhadap 50 sampel jawaban siswa menunjukkan nilai OUTFIT MNSQ sebesar 0,98, sehingga dapat disimpulkan bahwa terdapat kesesuaian antara siswa dengan Rasch model, kecuali siswa yang bernomor 04L, 18P, 07P, 44P, 39L, 43P dan 25P.

Parameter kedua *Outfit Z-standard* (OUTFIT ZSTD). Hasil analisis menunjukkan bahwa nilai *Outfit Z-standard* (OUTFIT ZSTD) yang diperoleh adalah sebesar 0,1. Hal ini menunjukkan bahwa data tersebut mempunyai kemungkinan nilai yang rasional.

Parameter yang ketiga adalah *Point measure correlation* (PTMEASURE). Hasil analisis jawaban dari 50 siswa menunjukkan bahwa nilai *Point measure correlation* (PTMEASURE) yang diperoleh berkisar antara 0,13 hingga 0,69. Hasil ini menunjukkan bahwa terdapat sebagian besar siswa yang termasuk kriteria fit $0,4 < \text{CORR} < 0,85$. Namun sebagian siswa tidak memenuhi kriteria ini, yaitu sejumlah 15 dari total 50 sampel peserta didik.

Validitas

Validitas adalah pengukuran harus valid, artinya instrumen pengukuran harus secara akurat mengukur apa yang sebenarnya ingin diukur. Validitas soal berdasarkan hasil analisis menggunakan pemodelan Rasch model dapat dilihat pada gambar 2. Analisis data menunjukkan bahwa terdapat sebagian soal yang valid dan beberapa soal yang tidak valid. Item soal yang valid berdasarkan hasil analisis ini adalah nomor, S2, S3, S4, S5, S9, S10, S12, S13, S14, S15, S19, S21, S23, S24, S25, S27, S29. Sementara soal-soal yang dinyatakan invalid adalah nomor S1, S6, S7, S8, S11, S16, S17, S18, S20, S22, S26, S28, S30. Beberapa item yang dinyatakan tidak valid menunjukkan jika soal tersebut seharusnya direvisi atau dieliminasi dan diganti dengan yang baru. Validitas suatu *instrument* tes ini dipengaruhi oleh berbagai faktor, salah satunya adalah *instrument* itu sendiri, serta pengguna yang melakukan pengukuran dan peserta didik yang diukur (Sugiyono, 2014).

Reliabilitas

Reliabilitas memberikan hasil yang konsisten dan dapat diandalkan dalam pengukuran berulang. Reliabilitas dapat dilihat berdasarkan besarnya nilai Alpha Cronbach (lihat Gambar 7.)

SUMMARY OF 50 MEASURED Person								
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	18.3	30.0	.68	.45	.99	.1	.98	.1
S.D.	4.2	.0	.84	.05	.18	.9	.40	.8
MAX.	27.0	30.0	2.77	.65	1.38	2.1	2.77	3.1
MIN.	11.0	30.0	-.68	.42	.55	-2.0	.34	-1.4
REAL RMSE	.47	TRUE SD	.69	SEPARATION	1.49	Person RELIABILITY	.69	
MODEL RMSE	.45	TRUE SD	.70	SEPARATION	1.55	Person RELIABILITY	.71	
S.E. OF Person MEAN = .12								
Person RAW SCORE-TO-MEASURE CORRELATION = .99								
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .69								

Gambar 7. Summary Analisis Rasch Model

Berdasarkan hasil analisis Rasch model diperoleh hasil nilai Alpha Cronbach sebesar 0,69. Nilai ini menunjukkan tingkat keajegan atau konsistensi hasil penilaian yang dilakukan menggunakan soal ujian akhir semester Biologi. Tingkat keandalan suatu *instrument* tes ini berkorelasi dengan semakin tingginya nilai reliabilitas hasil pengukuran (Alafi, Khadijah, dkk, 2020). Nilai Alpha Cronbach hasil perhitungan $0,69 < 0,7$, hal ini menunjukkan bahwa butir item soal belum memenuhi standar reliabilitas. Artinya soal belum tepat sasaran. Hal ini juga menunjukkan bahwa diperlukan perbaikan pada soal sehingga soal menjadi lebih reliabel, sehingga keajegan hasil pengukuran menggunakan item soal menjadi lebih reliabel atau dapat dipercaya. Soal yang memerlukan perbaikan atau revisi adalah soal kategori invalid yaitu nomor S1, S6, S7, S8, S11, S16, S17, S18, S20, S22, S26, S28, S30.

KESIMPULAN

Berdasarkan hasil penelitian dapat disimpulkan bahwa bahwa dari 30 butir soal, sejumlah 17 item termasuk kategori valid, sedangkan 13 item lainnya invalid. Nilai Alpha Cronbach berdasarkan hasil analisis adalah $0,69 < 0,7$; menunjukkan bahwa beberapa soal yang invalid tersebut tidak reliabel dan memerlukan perbaikan, sehingga keajegan atau konsistensi hasil pengukuran menggunakan item soal menjadi lebih dapat dipercaya. Hasil analisis juga menunjukkan bahwa mayoritas siswa memiliki kesesuaian dengan RASCH model yang dikembangkan, kecuali siswa nomor 04L, 18P, 07P, 44P, 39L, 43P dan 25P.

DAFTAR PUSTAKA

- Ademi, N. (2017). Programming and Implementation of Assessment. *Journal of Business and Financial Affairs*. <https://doi.org/10.4172/2167-0234.1000291>
- Akbaev, R., & Babichev, N. (2023). Tests as a control of knowledge and a key link in improving the educational process. *Rossiiskij Veterinarnyj Zhurnal*. <https://doi.org/10.32416/2500-4379-2023-1-7-9>
- Alavi, K., Isa, K., & Palpanadan, S. T. (2020). Application of Rasch Model on Resilience in Higher Education: An Examination of Validity and Reliability of Malaysian Academician Happiness Index (MAHI). *International Journal of Higher Education*, 9(4), 261-271. <https://doi.org/10.5430/ijhe.v9n4p261>
- Chaudhary, S., & Dey, N. (2013). Assessment in open and distance learning system (ODL): A Challenge. *Open Praxis*, 5(3), 207-216. <http://dx.doi.org/10.5944/openpraxis.5.3.65>
- Dombi, J., & Jónás, T. (2021). Likert Scale-Based Evaluations with Flexible Fuzzy Numbers. https://doi.org/10.1007/978-3-030-51949-0_5
- Efendi, R., Yondri, S., & Irawati, Y. (2019). Pengembangan E-Authentic Asessment Berbasis PBL untuk Meningkatkan Kompetensi Mahasiswa dalam Pembelajaran Jaringan Komputer. <https://doi.org/10.29207/RESTI.V3I3.1390>
- Fachri, M. (2018). Urgensi evaluasi pembelajaran dalam pendidikan. <https://doi.org/10.33650/EDURELIGIA.V2I1.320>
- Febriyanti, K., Qoiria, S., & Ananda, M. (2022). Evaluation of Learning to Improve The Efficiency of Student Assessment at The SD/MI. *Journal of Quality Assurance in Islamic Education*. <https://doi.org/10.47945/jqaie.v2i1.611>
- Guttsayt, E. (2020). Evaluation of the Effectiveness of Activities: Results and Problems. <https://doi.org/10.12737/1998-0701-2020-43-50>
- Hingorjo, M. R., & Jaleel, F. (2012). Analysis of one-best MCQs: The difficulty index, discrimination index and distractor efficiency. *Journal of the Pakistan Medical Association*, 62(2), 142-147.
- Kaushik, R., & Jayaram, B. (2016). Structural difficulty index: a reliable measure for modelability of protein tertiary structures. *Protein Engineering Design & Selection*. <https://doi.org/10.1093/PROTEIN/GZW025>
- Maba, W., Perdata, I. B. K., Astawa, I. N., & Mantra, I. B. N. (2018). Conducting assessment instrument models for teacher competence, teacher welfare as an effort to enhance education quality. *International Research Journal of Management, IT and Social Sciences*. <https://doi.org/10.21744/IRJMIS.V5I3.667>
- Magdalena, I., & Apriliyani, D. (2023). Pentingnya Peran Evaluasi dalam Pembelajaran di Sekolah Dasar. *MASALIQ*. <https://doi.org/10.58578/masaliq.v3i5.1381>
- Metsämuuronen, J. (2020). Generalized Discrimination Index. <https://doi.org/10.12973/IJEM.6.2.237>
- Olson, C. A., & Bakken, L. L. (2017). Evaluations of educational interventions: getting them published and increasing their impact. *Journal of Continuing Education in The Health Professions*. <https://doi.org/10.1002/CHP.21168>
- Oermann, M. H. (2018). End-of-Course Examinations: 10 Tips for Preparing Your Tests. *Nurse Educator*. <https://doi.org/10.1097/NNE.0000000000000511>
- Pospíšil, J., Pospíšilová, H., & Siarda Trochtová, L. (2022). The Catalogue of Leisure Activities: A New Structured Values and Content Based Instrument for Leisure Research Usable for Social Development and Community Planning. *Sustainability*. <https://doi.org/10.3390/su14052657>
- Rahma, N. A. A., Shamad, M. M. A., Idris, M. E. A., Elfaki, O. A., Elfakey, W. E. M., & Salih, K. M. A. (2017). Comparison in the quality of distractors in three and four options type of multiple choice questions. *Advances in Medical Education and Practice*. <https://doi.org/10.2147/AMEP.S128318>
- Rapono, M., Safrial, S., & Wijaya, C. (2019). Urgensi Penyusunan Tes Hasil Belajar: Upaya Menemukan Formulasi Tes Yang Baik dan Benar. <https://doi.org/10.24114/JUPIIS.V11I1.12227>

-
- Reotutar, M. A. C., Tactay, N. T., & Ridwan, M. (2020). *Achievement Test of Education Students in Assessment of Student Learning*. <https://doi.org/10.33258/BIRLE.V3I4.1330>
- Sudijono (2008). *Pengantar Evaluasi Pendidikan*. Jakarta: Rajagrafindo Persada.
- Sugiyono. (2014). *Statistika untuk penelitian*. Bandung : Alfabeta
- Vaiz, Osman, et al. (2021). A Review in Measurement and Assessment in Distance Education. *The Journal of New Horizon in Education*, 11 (4), 161 – 164.
- Weisburd, D., & Britt, C. (2014). *Measurement: The Basic Building Block of Research*. https://doi.org/10.1007/978-1-4614-9170-5_2
- Wibisono, S. (2018). Aplikasi Model Rasch Untuk Validasi Instrumen Pengukuran Fundamentalisme Agama Bagi Responden Muslim. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia (JP3I)*, 5(1). <https://doi.org/10.15408/jp3i.v5i1.9239>
- Wibowo, A., & Cholifah, T. (2018). Instrumen Tes Tematik Terpadu Kurikulum 2013 Berbasis PISA's Literacy bagi Siswa Sekolah Dasar. *JIPVA (Jurnal Pendidikan IPA Veteran)*, 2(2), 209-221. <https://doi.org/10.31331/jipva.v2i2.726>
- Wuang, Yee-Pay, et al. (2009). Rasch analysis of the Bruininks-Oseretsky Test of Motor Proficiency-Second Edition in intellectual disabilities. *Research in Developmental Disabilities*, 30, 1132-1144. <https://doi.org/10.1016/j.ridd.2009.03.003>
- Yusup, Febrianawati. (2018). Uji Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif. *Jurnal Tarbiyah: Jurnal Ilmiah Kependidikan*, 7 (1): 17 – 23. <https://doi.org/10.18592/tarbiyah.v7i1.2100>
- Ziborova, V., Katalnikova, S., Zagulova, D., & Prokofjeva, N. (2020). *Methods for evaluating the use of different approaches of organizing the process of teaching students*. <https://doi.org/10.17770/SIE2020VOL2.4978>
-