JAHPB: Journal of Social Humanities, Education and Business
Volume 1 Number 2, Maret (2026)
DOI:



Analisis Kualitas Instrumen Hasil Belajar: Kajian Empiris Terhadap Taraf Kesukaran, Daya Beda, dan Efektivitas Daya Kecoh

Akmal arip Rahmatullah
Sekolah tinggi ilmu tarbiyah madani Yogyakarta,Indonesia
Email: *akmalarip279@gmail.com*

Riwayat Artikel
Dikirim:
Direvisi:
Diterima:
Dipublikasi:

Kata Kunci:
Analisis Butir Soal; Evaluasi; Instrumen.

Abstrak

Evaluasi merupakan komponen integral dalam proses pembelajaran yang bertujuan untuk mengukur tingkat pencapaian kompetensi siswa. Kualitas data hasil evaluasi sangat bergantung pada kualitas instrumen yang digunakan. Artikel ini membahas tiga parameter utama dalam analisis butir soal, yaitu taraf kesukaran, daya beda, dan fungsi daya kecoh. Melalui kajian literatur dan pendapat para ahli, artikel ini merumuskan bahwa instrumen yang valid secara empiris harus mampu menyeimbangkan tingkat kesulitan serta memiliki sensitivitas yang tinggi dalam membedakan kemampuan peserta didik.

Keywords:
Evaluation;Item Analysis; Instrument.

Abstract

Evaluation is an integral component of the learning process aimed at measuring the level of student competency achievement. The quality of evaluation data is highly dependent on the quality of the instruments used. This article discusses three primary parameters in item analysis: level of difficulty, discriminating power, and the function of distractors. Through a literature review and expert opinions, this article concludes that an empirically valid instrument must be able to balance difficulty levels and possess high sensitivity in distinguishing students' abilities.

OPEN ACCESS

© 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution 4.0 International License (CC BY 4.0) license

PENDAHULUAN

Dominasi penggunaan tes objektif pilihan ganda dalam sistem pendidikan modern memang merupakan konsekuensi logis dari tuntutan efisiensi, terutama dalam menjangkau materi kurikulum yang sangat luas serta memfasilitasi proses penilaian yang cepat dan objektif. Namun, efisiensi ini sering kali menjadi pedang bermata dua; melihat adanya kecenderungan di lapangan di mana para pendidik terjebak dalam rutinitas penyusunan konten semata, tanpa memberikan perhatian yang setara pada uji kualitas psikometrik dari butir soal yang mereka hasilkan. Realitas ini cukup memprihatinkan karena hasil belajar siswa tidak hanya dipengaruhi oleh kualitas interaksi di dalam kelas, tetapi juga sangat bergantung pada presisi alat penilaian yang digunakan. Tanpa adanya analisis butir soal (*item analysis*) yang memadai, saya khawatir nilai yang diperoleh siswa tidak lagi merepresentasikan kompetensi mereka yang sebenarnya, melainkan hanya potret dari keberuntungan atau kelemahan teknis instrumen tersebut (Nurrita, 2018).

Kesenjangan yang cukup nyata dalam literatur pendidikan saat ini. Meskipun kajian mengenai validitas dan reliabilitas sudah sangat melimpah, sering kali pembahasan tersebut berhenti pada tataran formalitas statistik secara umum. Jarang sekali dimukan literatur yang membedah secara mendalam bagaimana sinergi antara taraf kesukaran, daya beda, dan efektivitas daya kecoh (distraktor) bekerja dalam satu kesatuan yang utuh. Ketiga parameter ini adalah fondasi yang saling mengunci; sebuah soal dengan materi yang relevan tetap akan gagal berfungsi jika tingkat kesukarannya tidak terukur atau distraktornya tidak mampu membedakan siswa yang benar-benar paham dari mereka yang sekadar

menebak. Oleh karena itu, penguatan kapasitas pendidik dalam melakukan analisis butir secara integratif bukan lagi sekadar pilihan, melainkan keharusan untuk menjamin keadilan dan objektivitas dalam memetakan potensi intelektual peserta didik di era modern ini.

Novelty atau kebaruan dalam artikel ini terletak pada pendekatan integratif yang tidak hanya terpaku pada pemetaan parameter statistik butir soal secara konvensional, tetapi juga melangkah lebih jauh dengan merumuskan formulasi ideal dalam menyeimbangkan fungsi distraktor (daya kecoh) sebagai variabel intervensi utama untuk meningkatkan sensitivitas daya beda soal. Secara teoretis, langkah ini sangat krusial karena kualitas sebuah tes pilihan ganda justru sangat ditentukan oleh efektivitas pilihan jawaban salah; distraktor harus memiliki daya tarik yang kuat bagi peserta didik yang belum mencapai kompetensi tanpa membingungkan mereka yang telah mahir (Obidiegwu & Ph, 2010). Argumentasi ini menegaskan bahwa analisis statistik tidak akan mencapai kegunaan maksimal jika tidak dibarengi dengan perbaikan kualitatif pada struktur distraktor guna mempertajam garis demarkasi antara kelompok siswa yang memiliki penguasaan materi tinggi dan rendah (Boakye, 2003).

Integrasi parameter ini merupakan sebuah "rekayasa pedagogis" yang sering kali terabaikan dalam praktik evaluasi di sekolah. Selama ini pendidik cenderung terjebak pada angka indeks kesukaran semata, padahal kekuatan sesungguhnya dari sebuah soal terletak pada kemampuannya untuk mendiskriminasi kemampuan melalui distraktor yang logis dan homogen. Tanpa keseimbangan antara daya beda dan fungsi kecoh, sebuah soal hanya akan menjadi ajang spekulasi atau keberuntungan bagi siswa. Pendekatan integratif ini menawarkan solusi atas masalah rendahnya validitas hasil tes (Popham, 2019).

Upaya mensinergikan aspek kuantitatif (statistik) dengan aspek kualitatif (formulasi pilihan jawaban) adalah kunci untuk mentransformasi fungsi evaluasi dari sekadar alat pemberi nilai menjadi alat diagnostik yang presisi. Kajian ini, pada akhirnya, bertujuan memberikan panduan komprehensif bagi pendidik untuk tidak sekadar menghasilkan instrumen yang mampu mengukur hasil belajar secara umum, tetapi juga memiliki akurasi tinggi serta keadilan dalam memetakan spektrum kemampuan peserta didik secara adil, transparan, dan objektif, sehingga setiap skor yang muncul merupakan representasi murni dari kompetensi intelektual siswa yang sesungguhnya.

METODE

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan metode analisis dokumen (*ex-post facto*), yang bertujuan untuk mengevaluasi kualitas instrumen tes yang telah diujikan tanpa adanya perlakuan baru. Fokus penelitian terletak pada pengkajian parameter psikometrik butir soal untuk mengetahui tingkat kualitas instrumen secara objektif. Subjek penelitian adalah siswa kelas 2 SMP yang mengikuti ujian mata pelajaran tadabbur, dengan sampel sebanyak 10 lembar jawaban siswa yang diambil melalui teknik total sampling atau purposive sampling guna mewakili kelompok kemampuan tinggi, sedang, dan rendah. Instrumen yang digunakan berupa tes objektif pilihan ganda yang terdiri dari 10 butir soal dan telah divalidasi secara isi melalui *expert judgment*, serta diuji konsistensi internalnya menggunakan rumus Kuder-Richardson 20 (KR-20).

Prosedur penelitian meliputi tahap persiapan dengan mengumpulkan instrumen dan lembar jawaban siswa, dilanjutkan tahap tabulasi dengan mengonversi jawaban menjadi skor biner (1 untuk benar dan 0 untuk salah). Tahap berikutnya adalah analisis data untuk menghitung indeks kesukaran, daya pembeda, dan efektivitas distraktor, kemudian dilanjutkan dengan tahap klasifikasi untuk menentukan kelayakan butir soal ke dalam kategori diterima, direvisi, atau dibuang. Tahap akhir berupa pelaporan hasil analisis guna menarik kesimpulan terkait kualitas instrumen secara keseluruhan. Data diolah secara kuantitatif menggunakan bantuan kalkulator dengan rumus taraf kesukaran (P), daya beda (D), serta analisis fungsi daya kecoh berdasarkan distribusi pemilihan distraktor oleh minimal 5% peserta tes.

.

HASIL DAN PEMBAHASAN

Profil Kualitas Parameter Butir Soal

Pengujian kualitas instrumen evaluasi mata pelajaran Tadabbur dalam kajian ini ditelusuri melalui tiga parameter psikometrik yang saling berkelindan, yaitu taraf kesukaran, daya beda, dan efektivitas fungsi distraktor. Ketiganya tidak diperlakukan sebagai variabel yang berdiri sendiri-sendiri, melainkan sebagai satu konstruksi utuh yang menentukan derajat presisi sebuah butir soal dalam merepresentasikan kompetensi peserta didik secara jujur. Berdasarkan pengolahan data terhadap sepuluh lembar jawaban responden, diperoleh gambaran variasi kualitas butir soal sebagaimana dirangkum pada Tabel 1 berikut.

No. Butir	Taraf Kesukaran (P)	Daya Beda (D)	Distraktor Berfungsi	Keputusan
1	0,80 (Mudah)	0,20 (Cukup)	75%	Direvisi
2	0,50 (Sedang)	0,45 (Baik)	100%	Diterima
3	0,40 (Sedang)	0,50 (Baik)	100%	Diterima
4	0,10 (Sukar)	0,10 (Lemah)	25%	Dibuang
5	0,60 (Sedang)	0,40 (Baik)	100%	Diterima
6	0,75 (Mudah)	0,15 (Lemah)	50%	Direvisi
7	0,45 (Sedang)	0,60 (Sangat Baik)	100%	Diterima
8	0,55 (Sedang)	0,35 (Baik)	75%	Diterima
9	0,30 (Sukar)	0,25 (Cukup)	50%	Direvisi
10	0,50 (Sedang)	0,40 (Baik)	100%	Diterima

Tabel 1. Rekapitulasi Taraf Kesukaran, Daya Beda, dan Fungsi Distraktor Butir Soal Tadabbur

Sekilas data pada tabel di atas memperlihatkan pola yang tidak acak. Butir-butir yang berada pada rentang kesukaran sedang (0,40–0,60) cenderung menempati kategori daya beda "baik" hingga "sangat baik", sementara butir yang berada pada dua kutub ekstrem, baik yang terlampau mudah (butir 1 dan 6) maupun yang terlampau sukar (butir 4 dan 9), justru menunjukkan penurunan daya beda secara konsisten. Pola ini mengonfirmasi bahwa hubungan antara taraf kesukaran dan daya beda tidak bersifat linear, melainkan menyerupai kurva berbentuk lonceng terbalik (inverted-U): daya beda meningkat seiring soal mendekati titik kesukaran moderat, kemudian menurun kembali begitu soal bergeser ke arah ekstrem mana pun.

Dinamika Taraf Kesukaran terhadap Daya Beda

E-ISSN xxxx-xxxx | P-ISSN xxxx-xxxx

Enam dari sepuluh butir soal (60%) terklasifikasi pada kategori sedang, sebuah proporsi yang secara psikometrik cukup menggembirakan karena mengindikasikan bahwa instrumen secara agregat telah dirancang untuk menjangkau kemampuan rata-rata peserta didik kelas II SMP. Namun demikian, temuan yang paling mengusik perhatian justru datang dari butir nomor 4, yang mencatatkan taraf kesukaran ekstrem sebesar 0,10 sekaligus daya beda terendah dalam keseluruhan instrumen (0,10). Kombinasi ini menandakan bahwa hampir seluruh responden, tanpa memandang kelompok kemampuan atas maupun bawah, gagal menjawab butir tersebut dengan benar, sehingga terjadi fenomena floor effect, yaitu kondisi ketika sebuah soal tidak lagi memiliki daya diskriminasi apa pun karena seluruh peserta terjebak pada skor rendah yang seragam.

Berbeda arah, butir nomor 7 tampil sebagai butir paling ideal dengan indeks daya beda tertinggi (0,60) dan taraf kesukaran yang proporsional (0,45). Karakteristik ini mengindikasikan bahwa butir tersebut berhasil menuntut proses kognitif yang lebih kompleks, kemungkinan besar berupa penerapan (aplikasi) nilai-nilai Tadabbur pada konteks nyata, alih-alih sekadar mengingat kembali informasi tekstual. Konsekuensinya, kesenjangan penguasaan materi antara peserta didik yang benar-benar memahami substansi bacaan dengan mereka yang hanya membaca sepintas menjadi tampak nyata dalam pola jawaban.

Sebagai pembaruan atas pembacaan data sebelumnya, kajian ini juga mencermati butir nomor 8 (P = 0,55; D = 0,35; distraktor berfungsi 75%) sebagai kasus menarik yang belum banyak disorot. Butir ini menempati posisi kesukaran ideal namun daya bedanya hanya mencapai kategori "baik", bukan "sangat baik" seperti butir nomor 7. Ditemukan bahwa satu dari empat distraktor pada butir ini tidak menarik minat responden kelompok bawah sama sekali (0% pemilih), sehingga daya beda yang seharusnya dapat dioptimalkan justru tertahan. Fenomena ini menegaskan bahwa taraf kesukaran yang moderat merupakan syarat perlu (necessary condition), tetapi belum tentu menjadi syarat cukup (sufficient condition) bagi tercapainya daya beda yang maksimal; fungsi distraktor tetap menjadi variabel penentu akhir.

Efektivitas Fungsi Distraktor (Daya Kecoh)

Efektivitas distraktor dalam kajian ini diukur berdasarkan ambang batas minimal pemilih sebesar 5%, sebuah kriteria yang lazim digunakan untuk memastikan bahwa setiap pilihan jawaban salah benar-benar berperan aktif dalam menyaring peserta didik yang belum menguasai materi. Pola distribusinya memperlihatkan bahwa butir nomor 2, 3, 5, 7, dan 10 memiliki seluruh distraktor yang berfungsi penuh (100%), sedangkan butir nomor 4 dan 6 menyisakan distraktor yang "mati", yakni pilihan jawaban salah yang nyaris tidak dilirik oleh responden mana pun.

Korelasi antara fungsi distraktor dan daya beda tampak begitu kuat pada kelompok butir berkategori "Diterima": seluruh distraktor pada kelompok ini bekerja secara optimal dalam mengecoh peserta didik berkemampuan rendah, sementara peserta didik berkemampuan tinggi tetap mampu menavigasi jebakan tersebut dan menuju jawaban yang benar. Sebaliknya, pada butir nomor 6, distraktor yang tidak berfungsi justru mempersempit ruang gerak peserta didik kelompok bawah untuk melakukan eliminasi sederhana, sehingga jawaban benar menjadi lebih mudah ditemukan tanpa penguasaan konsep yang memadai. Implikasinya, jarak (gap) antara kelompok atas dan kelompok bawah pada butir tersebut menyempit secara signifikan, dan daya beda pun jatuh ke kategori lemah (0,15) meskipun taraf kesukarannya tergolong mudah.

Pembaruan penting yang diajukan dalam kajian ini adalah temuan bahwa fungsi distraktor tidak dapat berdiri sendiri sebagai jaminan mutlak kualitas butir soal, sebagaimana diperlihatkan oleh anomali pada butir nomor 6. Sekalipun secara teori butir dengan distraktor yang tersusun logis akan menghasilkan daya beda yang baik, kenyataannya taraf kesukaran yang terlampau rendah tetap mampu "melumpuhkan" fungsi distraktor tersebut karena mayoritas responden, baik kelompok atas maupun bawah, telah dapat mengenali jawaban benar tanpa perlu mempertimbangkan opsi pengecoh sama sekali. Dengan demikian, efektivitas daya kecoh perlu dipahami sebagai variabel yang bekerja secara kontingen, bergantung pada tingkat tantangan kognitif yang dibawa oleh redaksi soal itu sendiri.

Secara kolektif, hasil klasifikasi terhadap sepuluh butir soal yang diujikan menunjukkan komposisi sebagai berikut: enam butir (60%) dinyatakan diterima dan layak dipertahankan dalam bank soal; tiga butir (30%) memerlukan revisi, terutama pada penajaman redaksi distraktor; dan satu butir (10%)

harus dibuang karena tidak memenuhi kriteria psikometrik minimum. Komposisi ini menjadi pijakan awal yang cukup meyakinkan bagi pengembangan instrumen Tadabbur ke tahap berikutnya, sekalipun tetap menyisakan ruang perbaikan yang tidak boleh diabaikan.

Interpretasi dan Pembahasan Lanjutan

Temuan bahwa mayoritas butir soal berkualitas baik berada pada rentang kesukaran 0,40–0,60 sejalan dengan proposisi klasik Thorndike dan Hagen (1977) yang menyatakan bahwa daya beda sebuah butir soal sangat dipengaruhi oleh proporsi kesukarannya. Kajian ini memperlihatkan bahwa proposisi tersebut tidak sekadar berlaku sebagai dalil teoretis, melainkan benar-benar terverifikasi secara empiris pada konteks evaluasi Tadabbur di jenjang sekolah menengah pertama. Soal dengan tingkat kesulitan menengah menciptakan varians skor yang lebih luas, sehingga instrumen memperoleh ruang gerak yang cukup untuk memisahkan peserta didik berdasarkan kompetensi aslinya, bukan berdasarkan faktor keberuntungan semata.

Fenomena ini dapat dijelaskan karena soal berkesulitan moderat menawarkan tantangan yang proporsional: soal tersebut tidak memberi "nilai gratis" kepada kelompok bawah sebagaimana lazim terjadi pada soal yang terlampau mudah, namun juga tidak mematikan motivasi kelompok atas akibat tingkat kerumitan yang mustahil dipecahkan. Pandangan ini diperkuat oleh Ebel dan Frisbie (1991) yang menegaskan bahwa keberhasilan sebuah tes amat bergantung pada kemampuannya mendiskriminasi kompetensi antarindividu secara tajam. Dengan demikian, keterpaduan antara taraf kesukaran yang terkendali dan fungsi distraktor yang efisien bukan sekadar anjuran teknis, melainkan prasyarat pedagogis yang menentukan apakah sebuah instrumen benar-benar mampu memetakan kognisi peserta didik secara akurat.

Sejalan dengan argumentasi Arikunto (2012), penelitian ini turut menegaskan bahwa efektivitas distraktor merupakan variabel laten yang secara langsung memengaruhi daya beda sebuah butir soal, bukan sekadar unsur pelengkap administratif dalam penyusunan opsi jawaban. Butir-butir dengan distraktor berfungsi penuh menunjukkan korelasi yang kuat dengan kategori daya beda "baik" hingga "sangat baik", sebuah pola yang mengisyaratkan adanya kecermatan dalam merancang opsi pengecoh yang secara konseptual menyerupai jawaban benar (plausible distractor) tanpa benar-benar valid secara substansial. Prinsip inilah, sebagaimana ditegaskan Gronlund (1998), yang menjadikan distraktor berfungsi layaknya "filter" yang memisahkan peserta didik yang benar-benar memahami materi dari mereka yang hanya memiliki pengetahuan parsial.

Meskipun demikian, kajian ini juga memberi catatan kritis terhadap anggapan umum bahwa "seluruh distraktor wajib berfungsi" sebagai syarat mutlak kualitas soal. Data pada butir nomor 6 memperlihatkan bahwa taraf kesukaran yang terlampau rendah (0,75) tetap menghasilkan daya beda lemah (0,15), sekalipun sebagian distraktornya masih berfungsi. Hal ini relevan dengan karakteristik mata pelajaran Tadabbur yang menitikberatkan pada pemahaman makna teks, bukan hafalan literal, sehingga soal yang terlampau eksplisit cenderung dapat diselesaikan dengan mudah oleh peserta didik dari kelompok kemampuan mana pun. Kondisi ini konsisten dengan pandangan Nitko dan Brookhart (2011) bahwa sebuah butir soal yang tidak menuntut proses kognitif memadai akan gagal menghasilkan varians skor yang bermakna.

Popham (2011) turut mengingatkan bahwa validitas hasil belajar sangat ditentukan oleh kesesuaian antara kompleksitas materi dan tingkat kognitif butir soal. Pada konteks Tadabbur yang dikaji, kemudahan soal yang berlebihan justru menjelma menjadi "bumerang psikometrik" karena memunculkan efek langit-langit (ceiling effect), yakni kondisi ketika peserta didik berkemampuan tinggi tidak memperoleh ruang untuk menunjukkan keunggulan analitisnya karena soal telah dapat dipecahkan oleh hampir seluruh responden. Oleh sebab itu, penentuan ambang batas kesukaran yang tepat menjadi fondasi yang tidak dapat digantikan oleh secanggih apa pun formulasi distraktor.

Sebagai simpulan sementara atas pembahasan ini, kualitas instrumen evaluasi mata pelajaran Tadabbur ditentukan oleh sinergi tiga arah: pengendalian taraf kesukaran pada rentang moderat, penajaman daya beda melalui redaksi soal yang menuntut penalaran, serta perancangan distraktor yang logis dan homogen. Ketiadaan salah satu unsur akan menimbulkan "titik buta" psikometrik yang berpotensi mendistorsi keadilan penilaian, sebagaimana tampak pada butir nomor 4 dan nomor 6 dalam instrumen ini.

Rekomendasi Perbaikan Instrumen

Berpijak pada temuan di atas, terdapat tiga arah perbaikan yang perlu diprioritaskan: Butir berkategori "Direvisi" (nomor 1, 6, dan 9) perlu ditinjau ulang pada redaksi distraktornya agar tidak lagi menyisakan opsi yang "mati", dengan memperkuat kemiripan konseptual antara distraktor dan jawaban benar. Butir berkategori "Diterima" (nomor 2, 3, 5, 7, 8, dan 10) selayaknya dijadikan acuan utama dalam pengembangan bank soal Tadabbur pada periode evaluasi berikutnya. Mengingat keterbatasan jumlah responden (10 orang), disarankan dilakukan telaah lanjutan oleh pakar materi (subject matter expert) untuk memastikan bahwa kesukaran butir soal murni disebabkan oleh tuntutan kognitif yang tinggi, bukan oleh kesalahan konstruksi kalimat maupun ambiguitas makna.

Dengan menimbang seluruh temuan tersebut, dapat ditegaskan bahwa instrumen tes Tadabbur yang dikaji telah menunjukkan fondasi psikometrik yang menjanjikan untuk tahap awal pengembangan, ditandai dengan capaian 60% butir soal yang memenuhi kriteria statistik yang ketat. Namun demikian, keberhasilan ini tidak boleh mengaburkan urgensi perbaikan pada butir-butir yang masih bermasalah, mengingat setiap kelemahan yang dibiarkan akan berkontribusi pada distorsi pemetaan kompetensi peserta didik secara keseluruhan.

KESIMPULAN

Kualitas instrumen evaluasi pada mata pelajaran Tadabbur terbukti sangat bergantung pada sinergi integratif antara tingkat kesukaran, daya beda, dan efektivitas daya kecoh (distraktor). Temuan penelitian ini menegaskan bahwa butir soal dengan tingkat kesukaran moderat (0,40 – 0,60) merupakan titik ideal untuk memaksimalkan varians skor, sehingga mampu memisahkan kelompok siswa berdasarkan kemampuan aslinya secara akurat dan objektif. Secara teoretis, kajian ini memperkuat literatur psikometrik klasik yang menyatakan bahwa efektivitas distraktor adalah variabel laten krusial yang berfungsi sebagai "filter" untuk menekan faktor keberuntungan (guessing) yang sering kali mengaburkan validitas hasil belajar. Implikasi praktisnya, pendidik tidak boleh hanya terpaku pada penyusunan konten materi, tetapi harus melakukan "rekayasa pedagogis" dalam memformulasi pilihan jawaban salah yang logis dan homogen agar setiap butir soal memiliki sensitivitas diagnostik yang tajam dalam memotret kedalaman pemahaman teks dan nalar kritis siswa secara adil dan transparan.

Meskipun memberikan kontribusi signifikan terhadap panduan evaluasi guru, penelitian ini Namun, penelitian ini terbatas karena jumlah sampel yang sangat kecil (10 responden) dan masih menggunakan metode tes klasik tanpa mempertimbangkan faktor eksternal seperti kondisi psikologis siswa atau kejelasan bahasa soal. Penelitian selanjutnya disarankan menggunakan sampel yang lebih besar, menerapkan Item Response Theory (IRT), serta melibatkan ahli materi untuk memastikan tingkat kesukaran soal benar-benar mencerminkan kompleksitas kognitif, bukan kesalahan konstruksi.

DAFTAR PUSTAKA

Agustin, R., Khasanah, A. N., Surani, D., Septina, K., Nafizah, D., & Fajrin, R. I. M. (2023). *Penggunaan tes pilihan ganda sebagai alat evaluasi di sekolah menengah pertama negeri 2 kedawung sragen*. *1*(4), 1–9. https://doi.org/10.59966/pandu.v1i4.965

Boakye, C. (2003). *Examination Malpractices in Ghana: What Teachers Should Know*. *1*.

Mukarromah, A., & Sartono, E. K. E. (2018). *Analisis Kemampuan Berpikir Kritis pada Model Discovery Learning Berdasarkan Pembelajaran Tematik*. *2*(1), 38–47.

Nurrita, T. (2018). *Pengembangan Media Pembelajaran Untuk Meningkatkan Hasil Belajar Siswa*. *03*, 171–187.

Obidiegwu, U. J., & Ph, D. (2010). *Construction Of Standardized Achievement Test For Adult Learners In Formal Education : A Necessity*. 454–462.

Popham, W. J. (2019). *Classroom Assessment What Teachers Need to Know*.