



Pengembangan *Computerized Adaptive Test (CAT)* Pada Materi Getaran, Gelombang dan Bunyi

Husnul Khotimah[✉], Budi Naini Mindyarto

Jurusan Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Semarang, Indonesia
 Gedung D7 Lt. 2, Kampus Sekaran Gunungpati, Semarang 50229

Info Artikel

Sejarah Artikel:

Diterima April 2021

Disetujui April 2021

Dipublikasikan Mei 2021

Keywords:

computerized adaptive test (cat), vibration, wave and sound

Abstrak

Permasalahan dalam evaluasi pembelajaran seperti *Paper Based Testing (PBT)* dan *Computer Based Testing (CBT)* yang mengabaikan kemampuan variasi individu menyebabkan informasi yang didapatkan kurang optimal. Para ahli telah mengembangkan alat evaluasi pembelajaran yaitu *Computerized Adaptive Test (CAT)* untuk mengatasi permasalahan tersebut. Pada sistem CAT, butir soal yang diberikan kepada peserta tes disesuaikan dengan tingkat kemampuan peserta tes. Tujuan penelitian ini untuk mengetahui karakteristik alat evaluasi CAT pada topik soal getaran, gelombang dan bunyi serta dapat mengetahui profil hasil belajar siswa materi getaran, gelombang dan bunyi dengan menggunakan alat evaluasi CAT. Penelitian ini menggunakan metode *Research and Development (R&D)* dengan model pengembangan 4-D yaitu *define, design, develop* dan *disseminate*. CAT yang dikembangkan berbasis *website* dan memiliki bank soal sebanyak 78 butir dengan topik soal getaran, gelombang dan bunyi. Hasil pengembangan yang didapatkan adalah Validitas instrumen mendapatkan skor validitas Aiken's *V* dengan rata-rata 0.93 yang melebihi validitas minimal sebesar 0,87 dan dinyatakan valid. Validitas ahli media mendapatkan skor 86% dengan kategori sangat baik. Metode analisis soal menggunakan *Item Response Theory (IRT)* model logistik satu parameter (1 PL). Tingkat kesukaran soal yang didapatkan yaitu kategori mudah sebanyak 17 butir, sedang sebanyak 30 butir dan sukar sebanyak 31 butir. Hasil penelitian didapatkan kelayakan media CAT sebesar 74% dengan kategori baik. Didapatkan profil hasil belajar siswa materi getaran gelombang dan bunyi menggunakan CAT dengan rata-rata nilai 47,91 dikategorikan sedang sedikit di bawah rata-rata dari nilai maksimal 100. Tingkat kemampuan peserta tes dapat di estimasi hanya dengan 11-23 butir soal yang mengakibatkan efisiensi waktu yang lebih baik dari PBT dan CBT. Terdapat hubungan antara skor menggunakan CAT dan skor tanpa menggunakan CAT secara positif dengan nilai korelasi r_{hitung} sebesar 0,600192, nilai ini dapat ditafsirkan dengan tingkat hubungan yang kuat, apabila skor tanpa CAT tinggi maka skor yang didapatkan dengan CAT juga tinggi, begitu sebaliknya.

Abstract

*Problems in learning evaluation such as Paper Based Testing (PBT) and Computer Based Testing (CBT) which ignore the ability of individual variations cause the information obtained to be less than optimal. Experts have developed a learning evaluation tool, namely the Computerized Adaptive Test (CAT) to overcome these problems. In the CAT system, the items given to test takers are adjusted to the test taker's ability level. The purpose of this study was to determine the characteristics of the CAT evaluation tool on the topic of vibration, wave and sound as well as to be able to determine the profile of student learning outcomes in vibration, wave and sound materials using CAT evaluation tools. This study uses the Research and Development (R&D) method with a 4-D development model, namely define, design, develop and disseminate. The CAT developed is website-based and has a 78 item question bank with the topic of vibration, waves and sound. The result of the development obtained is that the validity of the instrument gets Aiken's *V* validity score with an average of 0.93 which exceeds the minimum validity of 0.87 and is declared valid. The validity of the media expert got a score of 86% in the very good category. The method of question analysis uses the Item Response Theory (IRT) one-parameter logistic model (1 PL). The level of difficulty of the questions obtained is easy category as many as 17 items, medium as many as 30 items and difficult as many as 31 items. The results showed that the feasibility of CAT media was 74% in a good category. Obtained profiles of student learning outcomes material vibration waves and sound using CAT with an average value of 47.91 categorized as being slightly below the average of the maximum value of 100. The ability level of test takers can be estimated with only 11-23 items resulting in efficiency better time than PBT and CBT. There is a relationship between the score using CAT and the score without using CAT positively with a correlation value r_{count} of 0.600192, this value can be interpreted as a strong level of relationship, if the score without CAT is high then the score obtained with CAT is also high, and vice versa.*

©2021 Universitas Negeri Semarang

[✉] Alamat korespondensi:
 E-mail: ayatulhusnul@gmail.com

PENDAHULUAN

Evaluasi merupakan salah satu komponen yang sangat penting untuk mengetahui keefektifan pembelajaran oleh guru (Arifin, 2012). Evaluasi ini dilakukan dengan cara pemberian tes kepada pembelajar. Tes digunakan di sekolah seperti ulangan, ujian tengah semester, ujian semester bahkan ujian nasional. Tes adalah pemberian tugas dalam bentuk soal atau perintah lain yang harus dikerjakan oleh peserta didik (Asrul, Ananda, & Rosnita, 2015).

Di Indonesia saat ini masih menggunakan tes yang konvensional dengan menggunakan administrasi *Paper Based Test* (PBT) (Yulianto, 2008). Menurut Tedyyana & Danuri (2017) PBT merupakan tes tertulis dengan membagikan lembar soal pada kertas lalu siswa mengisi lembar jawaban yang telah disediakan oleh panitia. Menurut Setemen (2010) penggunaan PBT memiliki banyak kelemahan. Pertama, evaluasi PBT memerlukan waktu dan biaya yang cukup banyak untuk memproduksi instrumennya. Kedua, pemilihan butir tes dari bank soal cukup merepotkan dalam hal pemilihan dan memproduksinya. Ketiga, proses pemeriksaan evaluasi dengan instrumen tercetak cukup rumit, sehingga memerlukan waktu banyak dan cenderung membosankan. Keempat, proses pengolahan skor dan umpan balik kepada responden juga rumit, memerlukan waktu banyak sehingga membosankan. Kelima, secara psikologis evaluasi manual sering menimbulkan kecemasan pada peserta tes. Pengawas yang berada di sekitar, kondisi peserta tes yang lain, serta keadaan lingkungan sekitar sering membuat peserta tes merasa cemas.

Kelemahan PBT diatas sangat mendesak untuk diatasi. Maka para ahli menggunakan teknologi komputer untuk membantu mengurangi kelemahan dari PBT. Membuat tes dengan komputer serta bekerja sama dengan programmer komputer evaluasi akan semakin efisien, kecurangan tes akan semakin berkurang dan pemberian skor pada siswa akan lebih objektif dan konsisten (Hadi, 2013).

Pemanfaatan tes dengan bantuan teknologi komputer ini dinamakan *Computerized Based Testing* (CBT). Sistem CBT pada dasarnya sama dengan PBT, yaitu setiap peserta tes menerima seperangkat butir tes yang sama, karena butir soal tidak tercetak di selembaran kertas dimungkinkan tes akan muncul di komputer secara acak. Tes CBT ini akan mengurangi tingkat kecurangan dalam melakukan tes (Anggrenita, Muslim, & Saf, 2016).

Pemberian butir soal dengan jumlah butir yang sama serta desain yang sama kepada peserta tes akan mengakibatkan pengabaian terhadap variasi kemampuan peserta tes, selain itu adanya pemberian butir-butir soal yang terlalu mudah untuk beberapa peserta tes yang memiliki kemampuan yang tinggi serta pemberian butir-butir soal yang terlalu sulit untuk peserta tes yang memiliki kemampuan yang sangat rendah. Dari penggerjaan soal yang terlalu mudah bagi peserta tes yang pandai dan penggerjaan soal yang terlalu sulit bagi peserta tes yang kemampuannya rendah sesungguhnya kurang memberikan informasi mengenai kemampuan peserta tes. Hal ini mengakibatkan tes kurang efisien, kurang akurat dan kurang adil (Hambleton, Swaminathan, & Rogers, 1991).

Untuk meningkatkan efisiensi dan akurasi dalam mengukur kemampuan siswa maka para ahli telah mengembangkan ujian komputer yang bersifat adaptif yaitu *Computerized Adaptive Test* (CAT) (Nashihah, Purwanti, & Widodo, 2017). Menurut Lord (1980) *Adaptive* dimaksudkan bahwa butir soal yang diberikan kepada peserta tes disesuaikan dengan tingkat kemampuan peserta tes. CAT menggunakan analisis soal *Item Response Theory* (IRT), karena IRT dapat memperhitungkan kemungkinan orang memberikan jawaban dengan benar untuk setiap item soal.

CAT tidak hanya memindahkan soal dari PBT ke komputer seperti CBT, tetapi komputer telah diatur oleh program untuk menyeleksi soal yang akan ditampilkan ke siswa, lalu sistem akan memberi skor pada siswa. Setelah siswa menjawab butir soal

tersebut maka komputer akan memilihkan butir soal yang baru kepada siswa. Butir soal yang akan diberikan memberikan informasi tertinggi untuk siswa berdasarkan jawaban soal sebelumnya, proses sistem ini akan terus berlanjut sampai aturan berhenti telah tercapai. Dari proses sistem ini umumnya peserta tes akan menerima butir soal yang sesuai dengan kemampuan masing-masing peserta tes dan menghindari peserta tes mendapatkan soal yang terlalu sulit atau terlalu mudah(Santoso, 2010).

Dalam penelitian Santoso (2010) terdapat empat kelebihan CAT di banding PBT dan CBT. Pertama, tes CAT dapat memberikan item soal sesuai dengan kemampuan individu peserta tes sehingga dapat meningkatkan efisiensi, butir soal yang terlalu mudah atau sulit dapat dihindari sehingga panjang waktu tes dapat berkurang tanpa mengurangi tingkat presisi pengukuran. Kedua, karena CAT mengambil soal yang telah ter kalibrasi dan tersimpan secara elektronik maka keamanan tes dapat terjaga lebih baik. Ketiga, skor peserta tes dapat langsung diketahui, karena komputer secara langsung menskor dan mengestimasi kemampuan peserta tes segera setelah peserta tes menjawab item soal yang diberikan. Keempat, dengan memanfaatkan komputer makan format tampilan pada item soal dapat berupa animasi atau suara yang lebih menarik dari tes konvensional.

CAT sudah banyak digunakan untuk tes kepegawaian di banyak negara. Seperti GMAT (*Graduate Management Admission Test*) (Fahmi, Alkaff, & Bilfaqih, 2011). Jepang juga telah meluncurkan CAT secara online dengan nama J-CAT pada <http://www.j-cat.org/> (Nashihah, Purwanti, & Widodo, 2017).

Dalam perkembangan *Computerized Adaptive Test (CAT)* di Indonesia terbilang masih tertinggal dibandingkan dengan negara lain. Perkembangan CAT di Indonesia masih tahap uji coba dan tahap penelitian. Salah satu lembaga yang mengembangkan CAT adalah Puspendik pada tahun 2007 dibawah Kementerian Pendidikan Nasional. CAT ini berhasil diuji cobakan sekali via internet dan

kali kedua (2008 dan 2009) via internet juga (Bagus, 2010).

Menurut Elyas (2018) Sekarang ini perkembangan IPTEKS menunjukkan peningkatan yang sangat berarti bagi kepentingan dunia pendidikan. Selain tuntutan teknologi, pada tahun 2019-2020 terdapat tuntutan keadaan. Menularnya Covid-19 membuat pemberlakuan kebijakan belajar di rumah, dengan memanfaatkan teknologi informasi, seperti jaringan internet. Pendidik harus mengubah sistem pembelajaran yang awalnya belajar secara tatap muka, kini semua pembelajaran dilakukan secara daring (*online*). (Astini, 2020)

Dari latar belakang yang sudah diuraikan diatas, tuntutan penggunaan pembelajaran daring semakin meningkat dengan adanya tuntutan zaman dan keadaan. Dengan pembelajaran secara daring, pendidik juga dituntut untuk menggunakan evaluasi pembelajaran secara daring, dan salah satunya dapat menggunakan media *website*. Media ini dapat dibuka secara luas dan kapan pun, maka peneliti mengembangkan produk *Computerized Adaptive Test (CAT)* berbasis *website*. Untuk soal yang digunakan adalah materi getaran, gelombang dan bunyi. Peneliti memilih materi getaran, gelombang dan bunyi dikarenakan belum ada peneliti CAT yang menggunakan materi tersebut.

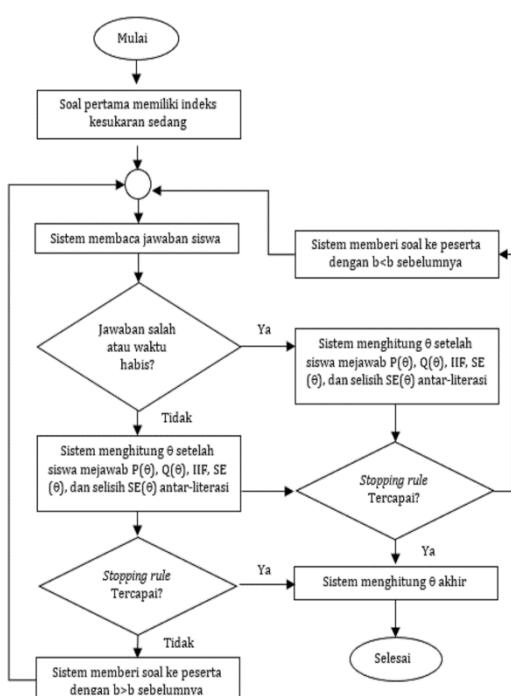
METODE

Lokasi penelitian dilakukan di SMP N 4 Demak, SMP N 1 Wonosalam dan SMP N 2 Wonosalam. Responden yang digunakan dalam penelitian yaitu siswa kelas IX IPA. Penelitian ini menggunakan metode penelitian dan pengembangan (*Research and Development*).

Metode penelitian dan pengembangan menurut Sugiyono (2016) adalah suatu cara ilmiah untuk meneliti, merancang, memproduksi dan menguji validitas produk yang dihasilkan. Menurut Thiagarajan (1974) langkah-langkah untuk penelitian dan pengembangan terdapat empat tahapan

yang disingkat 4D, yaitu *Define, Design, Development, and Dissemination*. Penelitian ini dilakukan sampai dengan *Development*. *Dissemination* tidak dilakukan karena keterbatasan waktu dan biaya oleh peneliti, sehingga peneliti melakukan tiga tahap, yaitu *Define, Design* dan *Development*. Teknik analisis data berupa validitas isi, tingkat kesukaran menggunakan IIRT, skala *Linkert*, korelasi *product moment*, dan analisis tingkat kognitif peserta tes.

HASIL DAN PEMBAHASAN



Bagan 1. Flowchart CAT

Computerized adaptive test membutuhkan *flowchart* untuk menjalankan sistem. Tampilan *flowchart* dapat dilihat pada bagan 1. Dari bagan tersebut maka diaplikasikan ke program CAT.

Pada gambar *Flowchart CAT*, mula-mula peserta tes diberikan soal awal yang memiliki nilai kesulitan sedang karena dianggap mempunyai tingkat kemampuan sedang. Jika soal dijawab dengan benar maka soal berikutnya peserta didik mendapatkan soal yang lebih sulit. Apabila soal sebelumnya dijawab salah, maka soal berikutnya

dimunculkan soal yang lebih mudah. Kemudian sistem menghitung nilai kemampuan (θ) setelah menjawab soal baru $P(\theta)$, $Q(\theta)$, IIF, $SE(\theta)$, dan selisih $SE(\theta)$ dan harga mutlak selisih kesalahan baku antar penyajian soal dihitung.

Rumus yang digunakan dalam buku Hadi (2013) yaitu:

$$\theta = b_i + \frac{1}{Da_i} \ln(0,5(1 + \sqrt{1 + 8c_i}))$$

$$P_i(\theta) = \frac{e^{(\theta-b_i)}}{1 + e^{(\theta-b_i)}}$$

$$Q_i(\theta) = 1 - P_i(\theta)$$

$$I_i(\theta) = P_i(\theta) Q_i(\theta)$$

$$SE(\theta) = \frac{1}{\sqrt{\sum_{i=1}^N I_i(\theta)}}$$

Penelitian dilaksanakan dalam tiga tahapan yaitu tahap pertama uji coba soal, tahap kedua uji coba media dan tahap ketiga penelitian. Uji coba soal memerlukan responden yang banyak untuk dapat dianalisis menggunakan teori respons butir. Dibutuhkan 200 responden untuk satu paket soal sehingga dalam penelitian ini dibutuhkan minimal 400 responden untuk dua paket soal, uji coba soal ini dilaksanakan pada kelas VIII di tiga sekolah menengah negeri di Demak yaitu SMP Negeri 1 Wonosalam, SMP Negeri 2 Wonosalam dan SMP Negeri 4 Demak. Tahap kedua uji coba media CAT dilaksanakan pada kelas IX di SMP Negeri 2 Wonosalam dengan responden 10 anak. Penelitian dilaksanakan pada kelas IX di dua sekolah negeri di Demak yaitu di SMP Negeri 2 Wonosalam dan SMP Negeri 4 Demak sebanyak 2 kelas.

Validitas Ahli Instrumen

Pada penelitian ini terdapat 82 butir soal yang diambil instrumennya di skripsi jurusan fisika UNNES di tahun sebelumnya. Terdapat 61 butir soal yang sudah valid (tidak ada revisi) dan 21 butir soal belum valid (terdapat revisi) menurut ahli dosen pembimbing. 61 butir soal tersebut dianggap valid karena telah melalui validitas pengujian sebelumnya, dan 21 butir soal dianggap belum valid karena masih terdapat kekurangan di dalamnya.

Hasil validitas menggunakan angket dengan empat kriteria yaitu: 1 = sangat tidak setuju, 2 = tidak setuju, 3 = setuju, dan 4 = sangat setuju. Hasil perhitungan menggunakan formula Aiken dengan jumlah 5 validator. Butir soal dikatakan valid apabila indeks validitas lebih besar dari 0,78. Hasil perhitungan didapatkan rentang indeks validitas isi untuk masing-masing butir soal adalah $0,78 \leq x \leq 1$. Indeks validitas isi yang didapatkan setiap butir soal menunjukkan butir soal yang valid. Dengan demikian ahli menyatakan bahwa instrumen tes sangat layak untuk diaplikasikan.

Validasi Ahli Media

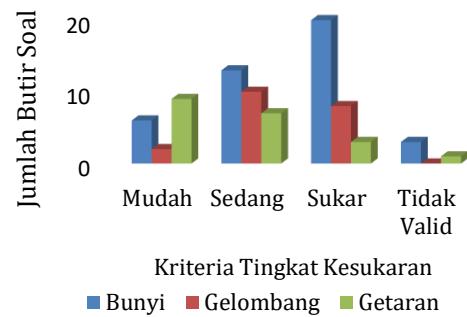
Produk yang dihasilkan dalam penelitian ini adalah aplikasi *computerized adaptive test* berbasis *website* sebagai alat evaluasi pembelajaran. Produk ini divalidasi para ahli sebanyak dua dosen. Peneliti memberikan angket penilaian dan tutorial penggunaan aplikasi kepada para ahli agar dapat menilai aplikasi *computerized adaptive test*. Para ahli mempunyai *username* dan *password* sebagai *admin* dan peserta.

Dari hasil yang didapatkan aplikasi *computerized adaptive test* mendapatkan skor lebih dari 80% untuk aspek nilai sistem, navigasi dan keandalan sistem sehingga dikategorikan sangat baik untuk diaplikasikan. Aspek tampilan soal mendapatkan skor 71% yang dikategorikan baik untuk diaplikasikan. Kelayakan *computerized adaptive test* dari para ahli rata-rata 86% yang tergolong sangat baik.

Tingkat Kesukaran Soal

Hasil analisis butir soal menggunakan IRT model 1 PL didapatkan tingkat kesukaran soal untuk masing-masing butir soal. Dari 2 paket soal terdapat 90 butir soal, di dalam 90 butir soal terdapat 8 butir soal jangkar sehingga total butir soal adalah 82 butir. Analisis untuk mendapatkan tingkat kesukaran menggunakan aplikasi BILOG-MG. Hasil yang didapatkan dari butir-butir soal yang memenuhi kriteria ideal yaitu memiliki tingkat kesukaran diantara $-3,0$ sampai $+3,0$.

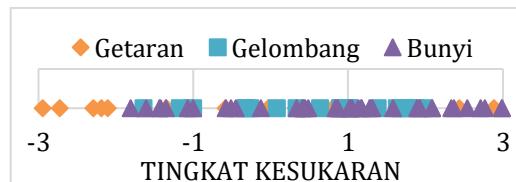
Soal yang memiliki keterangan tidak valid dibuang karena tidak masuk dalam kriteria. Berikut grafik hasil analisis kriteria tingkat kesukaran soal pada masing-masing topik soal.



Gambar 1. Hasil Analisis Kriteria

Tingkat Kesukaran Pada Topik Soal

Kriteria mudah memiliki jumlah tertinggi pada topik soal getaran. Sedangkan topik soal bunyi memiliki jumlah tertinggi pada kriteria sedang dan sukar. Hal ini juga dipengaruhi karena bunyi memiliki jumlah soal yang lebih banyak daripada getaran dan gelombang. Butir soal yang tidak valid berjumlah 4 butir. Butir soal yang dinyatakan valid berjumlah 78 butir. Butir soal yang ideal dapat dimasukkan ke aplikasi beserta kunci jawaban, tingkat kesukaran dan topik soal. Persebaran tingkat kesukaran dan topik soal dapat dilihat sebagai berikut.



Gambar 2. Persebaran Tingkat

Kesukaran Soal Materi Getaran, Gelombang dan Bunyi

Hasil yang didapatkan kriteria yang menunjukkan mudah sebanyak 17 butir dengan rentang tingkat kesukaran antara $-3 \leq x \leq -1$. Kriteria sedang berjumlah 30 butir dengan rentang tingkat kesukaran $-1 < x < 1$. Kriteria sukar berjumlah 31 butir dengan rentang $1 \leq x \leq 3$. Kriteria sukar dan sedang lebih banyak dari

kriteria mudah hal ini yang mengakibatkan persebaran soal di kriteria mudah dengan tingkat kesukaran $x \leq -1$ tidak banyak simbol kecil pada gambar 2.

UJI COBA MEDIA

Aplikasi *computerized adaptive test* berbasis *website* yang telah diuji oleh ahli media dan telah direvisi siap untuk di uji coba. Uji coba media ini dilaksanakan di SMP Negeri 2 Wonosalam sebanyak 10 anak. Dari hasil dan saran yang didapatkan dapat memperbaiki kualitas dari produk aplikasi *computerized adaptive test*.

Peserta tes menonton video tutorial dan mengerjakan demo tes getaran, gelombang dan bunyi agar lebih memahami sistem kerja dari aplikasi *computerized adaptive test*. Angket dimunculkan ketika peserta tes telah mengerjakan tes utama yaitu tes getaran, gelombang dan bunyi. Dari hasil yang didapatkan aplikasi *computerized adaptive test* mendapatkan skor lebih dari 90% untuk aspek nilai sistem, tampilan soal, navigasi dan keandalan sistem, sehingga dapat dikategorikan sangat baik untuk diaplikasikan.

Respons Siswa Menggunakan Computerized Adaptive Test

Aplikasi *computerized adaptive test* berbasis *website* dengan alamat <https://aplikasi-cat.my.id/> yang telah diuji oleh ahli dan di uji coba sebelumnya siap untuk di digunakan sebagai penelitian. Penelitian ini dilaksanakan di SMP Negeri 4 Demak dan SMP Negeri 1 Wonosalam, setiap sekolah diambil 1 kelas untuk penelitian. Setiap sekolah diambil 1 kelas untuk penelitian.

Sistem CAT yang dirancang berbasis web mempunyai kelebihannya yakni peserta dapat mengikuti tes dengan mudah dimana saja dengan syarat pengguna melakukan koneksi dengan server setelah mendapatkan hak akses. Penelitian ini dilaksanakan sejumlah dua kelas. Peserta tes mendapat kode peserta dan *password* dari peneliti. Kode peserta dan *password* berupa angka sehingga dapat

memudahkan siswa untuk masuk ke aplikasi. Setelah peserta tes berhasil masuk ke aplikasi, peserta tes menonton video tutorial dan mengerjakan demo tes getaran, gelombang dan bunyi agar lebih memahami sistem kerja dari aplikasi *computerized adaptive test*. Setiap butir soal diberi waktu selama 3 menit, apabila penggerjaan soal lebih dari tiga menit maka sistem akan menganggap soal tersebut salah.

Angket dimunculkan ketika peserta tes telah mengerjakan tes utama yaitu tes getaran, gelombang dan bunyi. Dari hasil yang didapatkan aplikasi *computerized adaptive test* mendapatkan skor lebih dari 70% untuk aspek nilai sistem, tampilan soal, navigasi dan keandalan sistem, sehingga dapat dikategorikan baik untuk diaplikasikan.

Secara umum aspek yang dinilai sudah baik untuk diaplikasikan. Aspek nilai sistem yang terdiri dari kemudahan pengoperasian sistem dan kemudahan memahami informasi pada *interface* sistem sudah baik. Aspek tampilan soal yang terdiri dari desain pada sistem dan tampilan soal peserta dinilai baik. Aspek Navigasi yang terdiri dari keberfungsian tombol navigasi, kejelasan ikon pada menu navigasi dan kemudahan menu navigasi untuk dipahami sudah baik untuk diaplikasikan. aspek keandalan sistem yang terdiri dari kecepatan sistem dalam menampilkan soal selanjutnya dan keberfungsian sistem ketika di *full screen* dinilai baik untuk diaplikasikan.

Peserta tes juga memberikan komentar untuk kemajuan dari aplikasi CAT ini yaitu; 1) sistem *computerized adaptive test* ini sangat baik, karena para siswa dapat mengetahui seberapa besar kemampuan mereka dan juga para siswa dapat mengerjakan soal dengan mandiri sesuai dengan total waktu yang sudah diberikan, 2) peserta tes sangat senang sekali bisa belajar menggunakan *computerized adaptive test* walaupun sedikit pusing karena harus memikirkan jawaban yang benar-benar matang, 3) sistem selalu beradaptasi dengan baik, 4) aplikasi CAT baik digunakan untuk daring, 5) peserta tes belajar banyak dari persoalan yang berada di CAT, 6) peserta tes berharap untuk di tingkatkan kualitas terbaik.

CAT memiliki beberapa kelemahan. Meskipun telah dilakukan dengan teknologi yang paling modern keberhasilan pelaksanaan CAT sangat tergantung dengan ketersediaan butir soal, banyaknya butir soal, kualitas butir soal dan bank soal. Jika butir yang tersedia kurang dari sisi jumlah, dan kualitas butir juga kurang baik, maka informasi yang diperoleh dari CAT juga kurang dan pemilihan butir menjadi terhambat.

Terkait dengan realitas bahwa data kebanyakan berdistribusi normal, maka peserta dengan kemampuan sedang berjumlah banyak sekali. Peserta-peserta ini dimungkinkan mengakses butir yang sama. Butir diakses oleh banyak peserta dinamai dengan *item exposure*. CAT disajikan dengan jaringan, apabila banyak peserta tes yang membuka aplikasi maka hal ini menyebabkan jaringan sibuk, sehingga butir menjadi sulit untuk dibuka.

Hasil Tes Dengan Computerized Adaptive Test

Berikut dipaparkan tabel 1 salah satu hasil tes getaran, gelombang dan bunyi dari salah satu peserta tes. Theta awal diberi skor 0 karena sistem belum mengetahui kemampuan masing-masing peserta tes. Karena kemampuan peserta belum diketahui maka sistem memberikan butir soal dengan tingkat kesukaran sedang.

Terlihat pada tabel bahwa butir pertama yang terpilih adalah butir soal dengan kode soal 236, memiliki tingkat kesukaran $b = -0,127$. Tingkat kesukaran pada butir soal awal bisa dikategorikan sebagai butir soal sedang, yang dipilih secara acak pada rentang tingkat kesukaran sedang (-0,5 sampai +0,5). Pada butir soal pertama ini memiliki skor 1 yang berarti peserta tes menjawab dengan benar dan theta jawab mengikuti b (tingkat kesukaran) pada butir soal tersebut.

Tabel 1. Hasil Tes Menggunakan CAT

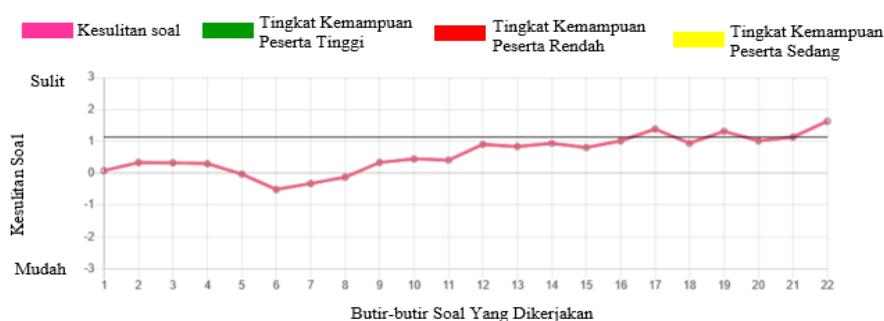
No	Kode Soal	Skor	B	θ	IIF		Selisih SE
					SE (θ)	2	
1	236	1	-0.127	0.127	0.25	2	2
2	258	0	-0.029	0.127	0.249	1.4151	0.5849
3	228	1	0.323	0.323	0.25	1.1552	0.2599
4	262	1	0.321	0.321	0.25	1.0003	0.1549
5	219	1	0.261	0.261	0.25	0.8946	0.1057
6	225	1	0.08	0.08	0.25	0.8167	0.078
7	272	1	0.41	0.41	0.25	0.7561	0.0606
8	266	1	0.899	0.899	0.25	0.7072	0.0488
9	287	1	1.015	1.015	0.25	0.6668	0.0405
10	222	1	1.385	1.385	0.25	0.6325	0.0342
11	280	0	1.586	1.385	0.247	0.6034	0.0292
12	248	0	1.172	1.385	0.247	0.5779	0.0254
13	263	1	1.015	1.015	0.25	0.5552	0.0227
14	286	1	1.054	1.054	0.25	0.535	0.0202
15	274	1	1.173	1.173	0.25	0.5168	0.0182
16	265	1	1.629	1.629	0.25	0.5004	0.0164
17	288	0	2.084	1.629	0.237	0.4861	0.0142
18	223	1	1.872	1.872	0.25	0.4724	0.0138
19	246	1	1.93	1.93	0.25	0.4597	0.0126
20	226	1	1.972	1.972	0.25	0.448	0.0117
21	235	0	2.367	1.972	0.24	0.4376	0.0104
22	277	1	2.084	2.084	0.25	0.4275	0.0101
23	259	0	2.084	2.084	0.242	0.4183	0.0092
Akhir Skor						2.084	
						84.73	

Tahap selanjutnya, karena pada item soal sebelumnya dijawab benar sistem memberikan kode soal 258. Kode soal 258 ini dipilih karena memiliki tingkat kesukaran yang lebih tinggi dari item soal sebelumnya. Tingkat kesukaran pada kode soal 258 sebesar $b = -0.029$. Selanjutnya butir soal kedua di respons salah, maka theta jawab sama dengan theta awal. Butir selanjutnya dengan kode 228 memiliki tingkat kesukaran lebih rendah dari sebelumnya sebesar $b = -0.321$. Proses berlangsung sampai *stopping rule* sistem tercapai. Pada tabel 4.35 selisih SE pada butir soal terakhir sebesar 0,0092 yang berarti telah mencapai salah satu aturan berhenti yaitu selisih SE kurang dari 0,01. Theta akhir yang didapatkan sebesar 2.084 atau dapat dikonversi menjadi skor 84,73.

Tabel 2. Statistik Soal

Tipe Soal	Jumlah
Bunyi	11
Gelombang	10
Getaran	2

Salah satu persyaratan dari CAT adalah peserta tes harus mengerjakan semua sub topik soal yang ada. Pada hasil tes getaran, gelombang dan bunyi ditampilkan sub topik soal yang telah dikerjakan. Sistem telah diatur agar semua sub topik soal harus dikerjakan oleh peserta tes. Pada hasil ini peserta tes telah mengerjakan sub topik soal bunyi sebanyak 11 soal, gelombang sebanyak 10 soal dan getaran sebanyak 2 soal.

**Gambar 3.** Grafik Kesulitan Soal

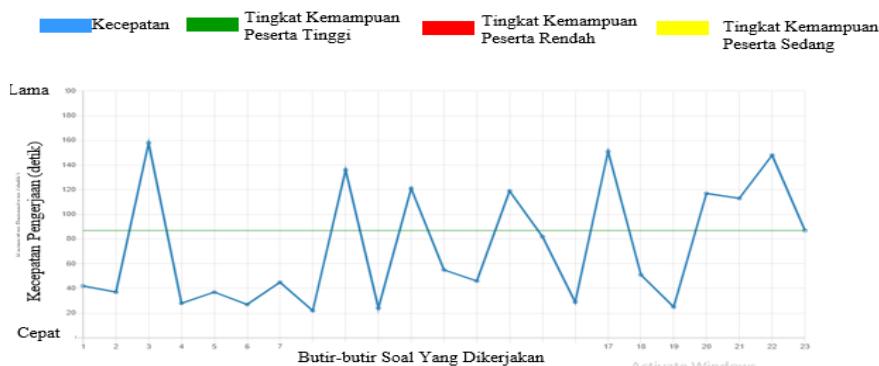
Grafik kesulitan soal ini agar mengetahui kenaikan dan penurunan tingkat kemampuan peserta, karena tingkat kemampuan peserta berbanding lurus dengan tingkat kesukaran soal. Pada butir soal nomor 1 menuju nomor dua digambar dengan garis menuju ke atas, ini menunjukkan peserta tes menjawab dengan

benar. Pada butir soal nomor 2 menuju nomor tiga digambarkan garis menuju ke bawah, ini menunjukkan peserta tes menjawab dengan salah. Pada butir-butir soal terakhir grafik menunjukkan kestabilan kemampuan peserta tes dan mencapai aturan berhenti sistem.

**Gambar 4.** Grafik Skor Tes

Grafik skor mengikuti theta atau tingkat kemampuan peserta apabila peserta tes menjawab benar. Apabila peserta tes menjawab salah maka theta jawab sama dengan theta awal, sehingga pada grafik skor tetap pada titik sebelumnya. Seperti titik pada

nomor 2 yang menjawab salah maka titik tersebut sama pada theta sebelumnya yaitu nomor satu. Apabila peserta menjawab benar maka titik pada grafik berubah seperti nomor 5 ke nomor 6.



Gambar 5. Grafik Kecepatan Penggerjaan Soal

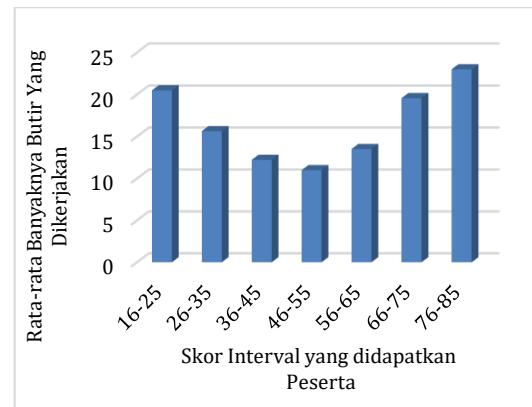
Selain grafik kesulitan soal dan grafik skor, sistem telah diatur agar memberikan grafik kecepatan penggerjaan. Manfaat dari grafik ini agar mengetahui peserta tes yang serius mengerjakan soal. Peserta tes yang tidak serius atau asal-asalan dalam penggerjaan soal akan tergambar pada grafik tersebut.

terdiri dari tingkatan ranah kognitif C1 sampai C6. Dari hasil yang didapatkan kemampuan kognitif dari peserta tes masih cukup baik, hal ini dapat menjadi evaluasi para pendidik agar mengevaluasi cara mengajar di kelas, selain itu peserta tes dapat mengevaluasi diri sendiri agar semakin giat belajar untuk meningkatkan kemampuan kognitif.



Gambar 5. Grafik Kategori Kemampuan Kognitif Peserta Tes

Kemampuan kognitif peserta tes dalam materi getaran, gelombang dan bunyi yang paling banyak dalam kategori cukup baik sebanyak 45 responden, dengan nilai antara 40 – 61. Peserta tes yang mendapatkan kategori sangat baik hanya satu anak, sedangkan kemampuan sangat rendah juga didapatkan satu anak. Kemampuan kognitif ini melibatkan kemampuan pengetahuan, pemahaman penerapan, sintesa dan evaluasi. Hal ini dikarenakan kumpulan bank soal

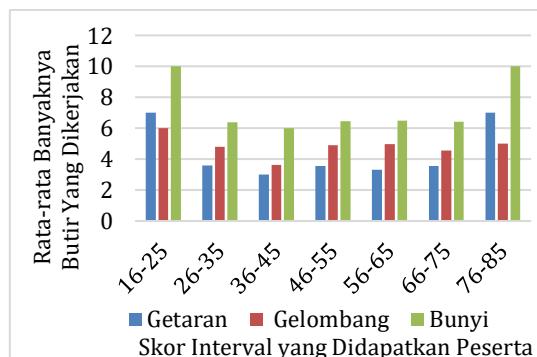


Gambar 6. Grafik Rata-rata Banyaknya Butir Soal yang Dikerjakan

Berdasarkan hasil tes getaran, gelombang dan bunyi diperoleh bahwa panjang tes (banyaknya butir soal yang dikerjakan) menggunakan aplikasi CAT setiap skor yang didapatkan dapat dilihat pada

gambar 6. Terlihat banyaknya butir yang diperlukan untuk mengestimasi tingkat kemampuan peserta tes berkisar antara 11 sampai 23 butir soal untuk rentang skor 16 – 85. Hal ini menunjukkan bahwa dengan rancangan tes adaptif tingkat kemampuan peserta tes sudah dapat di estimasi kemampuannya dengan menyajikan butir soal yang lebih sedikit atau setengah dari item soal yang dibutuhkan PBT dan CBT, karena dengan tes PBT atau pun CBT dibutuhkan soal sebanyak 40-45 untuk mengestimasi kemampuan peserta tes dalam ujian, seperti ujian nasional atau ujian sekolah.

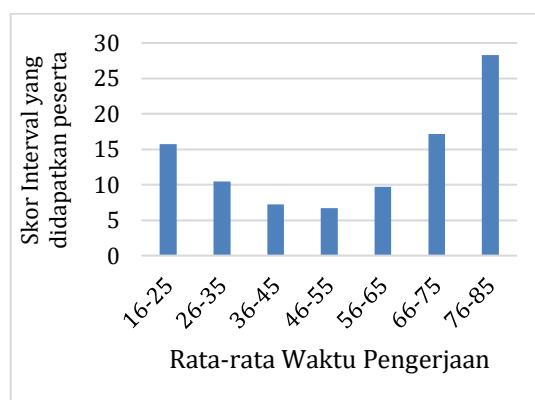
Terlihat pada grafik bahwa untuk tingkat kemampuan ekstrem (rendah maupun tinggi) dibutuhkan jumlah butir (panjang tes) yang lebih banyak untuk skor rendah 16 dibutuhkan 21 butir soal sedangkan skor paling tinggi 85 dibutuhkan 23 butir soal. Banyaknya butir soal dibawah skor 45-55 lebih sedikit daripada skor 45-55 ke atas, hal ini dikarenakan bank soal yang digunakan kurang menyediakan butir soal dengan tingkat kesukaran mudah untuk peserta dengan kemampuan rendah. Selain itu jumlah butir yang dikerjakan oleh peserta tes rata-rata terendah adalah 11, hal ini dikarenakan jumlah bank soal yang hanya tersedia 78 butir soal. CAT akan lebih maksimal apabila terdapat bank soal yang tersedia minimal 100 butir soal (Embretson & Reise, 2000).



Gambar 7. Banyaknya Butir Soal yang Dikerjakan Berdasarkan Sub Topik

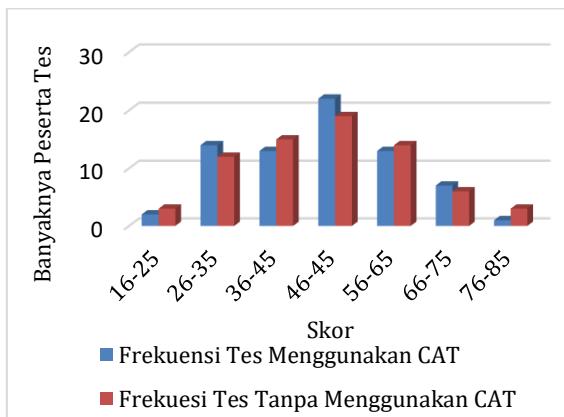
Butir soal yang diberikan oleh peserta tes harus menyajikan semua sub topik. Jumlah soal pada sub topik getaran sebanyak

19 butir, gelombang sebanyak 20 butir dan bunyi sebanyak 39 butir. Dari Gambar 7 hasil yang didapatkan sub topik bunyi lebih banyak di berikan kepada peserta tes, hal ini dikarenakan bank soal sub topik bunyi lebih banyak butir soal daripada sub topik lainnya. Dari rata-rata yang didapatkan butir soal getaran sebanyak 4 butir, gelombang sebanyak 5 butir dan bunyi sebanyak 7 butir. Dari hasil yang didapatkan antara butir soal yang diberikan kepada peserta tes menunjukkan keseimbangan konten pada aplikasi ini kurang baik, Karena jumlah soal pada sub topik yang diberikan kepada siswa kurang seimbang.



Gambar 8. Rata-rata Waktu Penggerjaan yang Dibutuhkan Oleh Siswa

Penyajian butir soal menggunakan CAT membutuhkan lebih sedikit dari tes yang lain mengakibatkan waktu yang lebih singkat untuk penggerjaan tes yang dilakukan. Pada Grafik 8 waktu yang dibutuhkan untuk mengerjakan soal selama 7-28 menit saja, rata-rata waktu penggerjaan selama 14 menit. Desain CAT lebih efisien waktu daripada PBT atau CBT, karena dengan desain PBT dan CBT peserta tes harus menyelesaikan semua butir soal yang disajikan selama 1,5 jam sampai 2 jam penggerjaan. Dengan model CAT maka waktu penggerjaan hanya membutuhkan setengah atau tiga perempat dari waktu penggerjaan menggunakan PBT atau CBT, sehingga CAT lebih efisien waktu dalam penggerjaan tes.



Gambar 9. Grafik Skor Peserta Tes Menggunakan CAT dan Tanpa CAT

Berdasarkan hasil tes getaran, gelombang dan bunyi menggunakan aplikasi CAT diperoleh skor peserta tes. Selain skor menggunakan CAT, peneliti juga mendapatkan skor ulangan getaran, gelombang dan bunyi tanpa menggunakan CAT dari guru mata pelajaran ilmu pengetahuan alam. Perbandingan skor yang didapatkan dapat dilihat pada gambar 9. Terlihat pada grafik selisih antara frekuensi tes menggunakan CAT dan frekuensi tes tanpa menggunakan CAT sangat sedikit yaitu selisih 1 sampai 3 peserta tes. Rata-rata skor yang didapatkan dari tes getaran, gelombang dan bunyi menggunakan CAT adalah 47,91, sedangkan skor ulangan tes getaran, gelombang dan bunyi tanpa menggunakan CAT memiliki rata-rata 48,33. Dari rata-rata nilai tersebut antara tes menggunakan CAT ataupun tidak memiliki selisih 0,42 sehingga dapat disimpulkan CAT dapat mengestimasi kemampuan peserta tes dengan dibuktikan skor yang didapatkan tes menggunakan CAT atau tidak memiliki skor yang hampir sama.

Interval nilai tes antara menggunakan CAT dan tanpa menggunakan CAT sangat bervariasi. Rata-rata interval nilai tes yang didapatkan adalah 10%. Interval yang didapatkan cukup tinggi, ini dikarenakan model yang digunakan pada CAT adalah 1PL yang hanya memperhatikan kesukaran soal. Model 1 PL ini tidak memperhatikan daya beda dan tebakan semu butir. Semakin banyak parameter yang digunakan maka program

semakin teliti dalam mengestimasi kemampuan peserta tes.

Skor yang didapatkan menggunakan aplikasi CAT dan skor yang didapatkan tanpa aplikasi CAT dapat dianalisis menggunakan korelasi *product moment*. Teknik ini bertujuan agar mengetahui hubungan antara kedua skor yang didapatkan dengan berbeda penerapan tes. Hipotesis nol (H_0) yaitu tidak ada hubungan antara skor menggunakan CAT dan skor tanpa menggunakan CAT, sedangkan hipotesis alternatif (H_a) yaitu terdapat hubungan antara skor menggunakan CAT dan skor tanpa menggunakan CAT.

Hasil yang didapatkan r_{hitung} adalah 0,600192, korelasi yang didapatkan adalah positif antara skor menggunakan CAT dan skor tanpa menggunakan CAT. Hal ini menunjukkan semakin tinggi skor yang didapatkan oleh CAT maka semakin tinggi pula skor yang didapatkan tanpa menggunakan CAT. Taraf kesalahan yang digunakan adalah 5% sehingga taraf kepercayaannya 95% yang didapatkan oleh r_{tabel} sebesar 0,227 dengan $N = 72$. Ternyata r_{hitung} lebih besar dari nilai r_{tabel} , sehingga H_0 ditolak dan H_a diterima. Jadi kesimpulannya yaitu ada hubungan positif dan nilai koefisien korelasi antara skor yang menggunakan CAT dan skor tanpa menggunakan CAT sebesar 0,600192.

Dari hasil r_{hitung} sebesar 0,600192 dapat ditafsirkan dengan tingkat hubungan yang kuat antara dua hasil skor tersebut, apabila skor tanpa CAT tinggi maka skor yang didapatkan dengan CAT juga tinggi, begitu sebaliknya. Koefisien yang didapatkan masih kurang sempurna, hal ini disebabkan model yang diadaptasi oleh CAT ini hanya menggunakan model 1 PL yaitu tingkat kesukaran butir soal, CAT semakin teliti apabila aplikasi menggunakan lebih banyak parameter lainnya seperti daya beda dan tebakan semu. Selain itu bank soal yang digunakan masih sedikit hanya 78 butir soal, apabila bank soal yang terdapat pada CAT semakin banyak maka ketelitian CAT semakin tinggi untuk mengestimasi kemampuan peserta tes.

SIMPULAN

Produk *computerized adaptive test* berbasis *website* memiliki bank soal materi Getaran, gelombang dan bunyi sebanyak 78 butir. Validitas isi Aiken's V mendapatkan skor 0,93 yang melebihi validitas minimal sebesar 0,87 dan dinyatakan valid. Analisis soal menggunakan *item response theory* model 1 PL dan didapatkan hasil tingkat kesukaran soal yaitu kategori mudah sebanyak 17 butir, sedang sebanyak 30 butir dan sukar sebanyak 31 butir. Kelayakan media CAT sebesar 74% dengan kategori baik untuk diaplikasikan. Hubungan antara skor menggunakan CAT dan skor tanpa menggunakan CAT dengan nilai korelasi r_{hitung} sebesar 0,600192 dengan kategori kuat, apabila skor tanpa CAT tinggi maka skor yang didapatkan dengan CAT juga tinggi, begitu sebaliknya. Tingkat kemampuan peserta tes dapat di estimasi hanya dengan 11-23 butir soal yang mengakibatkan efisiensi waktu yang lebih baik dari PBT dan CBT.

Produk akhir adalah aplikasi *computerized adaptive test* berbasis *website* dengan topik soal Getaran, gelombang dan bunyi yang mengharuskan peserta tes mengerjakan semua sub topik soal, sub topik soal memiliki perbandingan jumlah butir sebesar 1:1:2. Rata-rata butir soal yang dikerjakan peserta tes berdasarkan sub topik yaitu getaran sebanyak 4 butir, gelombang sebanyak 5 butir dan bunyi sebanyak 7 butir. Rata-rata waktu yang dibutuhkan untuk mengerjakan soal selama 14 menit. Tes CAT dapat memberikan item soal sesuai dengan kemampuan individu peserta tes, hal ini dapat menghindari butir soal yang terlalu mudah atau sulit sehingga penggerjaan lebih efisien waktu dan penyajian jumlah butir soal. Didapatkan profil hasil belajar siswa materi getaran gelombang dan bunyi menggunakan CAT dengan rata-rata nilai 47,91 dikategorikan sedang sedikit dibawah rata-rata dari skor maksimal 100.

DAFTAR PUSTAKA

- Anggrenita, L., Muslim, I., & Saf, M. (2016). Pengembangan Computerized Adaptive Test (CAT) Menggunakan Algoritma Fuzzy dengan Metode Tsukamoto Berbasis Web. *Jurnal Aksara Komputer Terapan*, 250-254.
- Arifin, Z. (2012). *Evaluasi Pembelajaran*. Jakarta: Direktorat Jenderal Pendidikan Islam Kementerian Agama RI.
- Asrul, Ananda, R., & Rosnita. (2015). *Evaluasi Pembelajaran*. Bandung: Perdana Mulya Sarana.
- Astini, N. K. (2020). Tantangan dan Peluang Pemanfaatan Teknologi Informasi Dalam Pembelajaran Online Masa Covid-19. *Jurnal Ilmu Pendidikan*, 241-255.
- Bagus, H. C. (2010). Sosialisasi Aplikasi Program CAT. Dalam Acara Uji Coba Aplikasi CAT Puspendik. Jakarta.
- Elyas, A. H. (2018). Penggunaan Model Pembelajaran E-Learning Dalam Meningkatkan Kualitas Pembelajaran. *Jurnal Warta*.
- Embretson, S., & Reise, S. (2000). *Item Response Theory for Psychologist*. New Jersey: Lawrence Elbaum Associated .
- Fahmi, M., Alkaff, A., & Bilfaqih, Y. (2011). Desain dan Implementasi Computerized Adaptive Test Berbasis Item Response Theory Pada Learning Management System Moodle. *Jurusan Teknik Elektro, Fakultas teknologi Industri, Institut Sepuluh November*, 1-7.
- Hadi, S. (2013). *Pengembangan Computerized Adaptive Test Berbasis Web*. (M. Romadhon, Ed.) Yogyakarta: Aswaja Pressindo.
- Hambleton, R., Swaminathan, H., & Rogers, H. (1991). *Fundamental of Item Response Theory*. Newbury Park, CA: Sage Publication Inc.

- Lord, F. (1980). *Applications of Item Respons Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Nashihah, K., Purwanti, D., & Widodo, D. (2017). Implementasi Sistem Penilaian Hasil Belajar Siswa Menggunakan Computer Adaptive Test (CAT). *Journal Of Economic Education*, 1-6.
- Santoso, A. (2010). Efisiensi dan Akurasi Computerized Adaptive Testing Pada Sistem Ujian Akhir Semester Universitas Terbuka. *Jurnal Pendidikan*, 11, 1-9.
- (2010). Pengembangan Computerized Adaptive Testing Untuk Mengukur Hasil Belajar Mahasiswa Universitas Terbuka. *Jurnal Penelitian dan Evaluasi Penelitian*, 62-83.
- (2010). Strategi Pemilihan Butir Alternatif Pada Tes Adaptif Untuk Meningkatkan Keamanan Tes. *Seminar Nasional FMIPA-UT*.
- Setemen, K. (2010). Pengembangan Evaluasi Pembelajaran Online. *Jurnal Pendidikan dan Pengajaran Universitas Ganeshaa*, 43, 207-214.
- Sugiyono. (2016). *Metode Penelitian dan Pengembangan (Research and Development/R&D)*. (S. Suryandari, Ed.) Bandung: Alfabeta.
- Tedyyana, A., & Danuri. (2017, November). Computer Based Test Untuk Seleksi Masuk Politeknik Negeri Bengkalis. *Jurnal Teknologi Informasi & Komunikasi Digital Zone*, 8, 114-124.
- Thiagarajan, S., etc. (1974). *Instructional Development for Training Teachers of Exceptional Children*. Washinton DC: National Center for Improvement Educational System.
- Yulianto, A. (2008). Computerized Adaptive Test : Pemanfaatan Psikologis dari Komputer Untuk Pengetasan Psikologi. *Prosiding Konferensi Nasional Sistem Informasi 2008: Indonesia Modern dalam Budaya Informasi*, 7-12.