

Data Driven Decision Making dalam Penentuan Kelayakan Soal Ujian Menggunakan Algoritma Fuzzy

Yudhi S. Purwanto^{1)}; Hendra Jatnika¹*

1. Program Studi Teknik Informatika, Fakultas Telematika Energi, Institut Teknologi PLN, Jl. Lingkar Luar Barat, Duri Kosambi, Cengkareng, Jakarta Barat, DKI Jakarta 11750 Indonesia

**)Email: y.purwanto@itpln.ac.id*

Received: 25 September 2024 | Accepted: 09 Desember 2024 | Published: 10 Januari 2025

ABSTRACT

The need for evident-based education makes some steps that must be taken by educational institutions. In relation to improving learning and student achievement, data driven decision making (DDDM) can be utilized to find out various things, such as the level of knowledge and skills of students at a personal level, one of which is through examinations. This research was conducted to determine the eligibility standards of college-level English exam questions using DDDM and fuzzy algorithms. 20 first-year students were tested to see the quality of the questions based on the results of the exam. The results are the level of difficulty of the question 0.611, Discrimination Power 20.2%, Validity produces 19 valid questions and 31 invalid questions, and reliability results are reliable. The conclusion is that DDDM method can be used to calculate the feasibility level of questions so that it can provide input on the development of exam questions standardization, especially for English. Further research should be conducted with larger data in several locations with different participants, thus the standard of questions' eligibility can be more accurate. The use of different algorithms or calculation models is also recommended so that it can provide benefits to this field.

Keywords: *DDDM, standardized test question eligibility, Fuzzy algorithm, English, Higher Education.*

ABSTRAK

Kebutuhan akan evident-based education membuat beberapa langkah yang harus diambil oleh pihak institusi Pendidikan. Dalam kaitannya dengan peningkatan pembelajaran dan prestasi peserta didik, data driven decision making (DDDM) dapat dimanfaatkan untuk dapat mengetahui berbagai hal, seperti tingkat pengetahuan dan keterampilan siswa dalam tataran personal, salah satunya melalui ujian. Penelitian ini dilakukan untuk mengetahui standar kelayakan soal ujian Bahasa Inggris Tingkat perguruan tinggi menggunakan DDDM dan algoritma fuzzy. 20 orang mahasiswa tingkat 1 diuji untuk melihat kualitas soal berdasarkan hasil ujian yang dilakukan. Hasilnya yaitu tingkat kesukaran soal 0.611 (sedang), Daya Beda 20,2% (cukup), Validitas menghasilkan 19 soal valid dan 31 soal tidak valid. Dalam perhitungan Reliabilitas didapatkan hasil yang reliable. Kesimpulan dari penelitian ini adalah bahwa metode DDDM dapat digunakan untuk menghitung tingkat kelayakan soal sehingga dapat memberikan masukan-masukan pada pengembangan standarisasi soal ujian, khususnya untuk bidang Bahasa Inggris. Penelitian selanjutnya sebaiknya dilakukan dengan data yang lebih besar dan pada beberapa lokasi dengan tingkat pengetahuan peserta yang berbeda-beda, sehingga standar kelayakan soal, khususnya bagian validitas dan reliabilitas juga dapat lebih akurat. Penggunaan algoritma atau model perhitungan yang berbeda juga disarankan sehingga dapat memberikan manfaat pada pengembangan bidang ini.

Kata kunci: *DDDM, standar kelayakan soal ujian, algoritma Fuzzy, Bahasa Inggris, Perguruan Tinggi*

1. PENDAHULUAN

Data driven decision making (DDDM) telah lama dikembangkan dalam dunia manajemen. Dalam perkembangannya, DDDM telah diakui sebagai sebuah alat yang mampu membantu pihak manajemen dalam memberikan berbagai Keputusan. Hal ini tentunya didukung oleh keberadaan bukti-bukti yang diolah sedemikian rupa, disesuaikan dengan kebutuhan dari manajemennya [1]. Ditambah lagi, perkembangan teknologi saat ini yang memungkinkan perhitungan dan manfaat DDDM dilakukan melalui perhitungan algoritma *machine learning* atau bahkan AI, sehingga dapat menyajikan hasil yang cepat, tepat dan akurat.

Manfaat DDDM juga merambah ranah Pendidikan saat ini. Pergeseran paradigma Pendidikan menjadi *evident-based education* membuat peran DDDM menjadi penting dalam berbagai bidang Pendidikan, mulai dari tingkat para pengambil Keputusan, sampai pada para guru/dosen di kelas [2]. Karena sifatnya yang berbasis database, DDDM juga dapat dimanfaatkan untuk tujuan yang beragam, dari mulai peningkatan inovasi dan pengembangan institusi Pendidikan secara berkesinambungan, mengidentifikasi ketidaksetaraan, baik di bidang kesetaraan gender maupun perolehan pendidikan, peningkatan akuntabilitas dan transparansi institusi Pendidikan, sampai pada peningkatan pembelajaran dan prestasi peserta didik [3].

Dalam kaitannya dengan peningkatan pembelajaran dan prestasi peserta didik, DDDM dapat dimanfaatkan untuk dapat mengetahui pola pembelajaran, pola belajar, mengidentifikasi dan mengatasi masalah secara cepat, dan mengetahui tingkat pengetahuan dan keterampilan siswa dalam tataran personal [4]. Tingkat pengetahuan dan keterampilan siswa dapat membuat guru untuk memformulasikan strategi pembelajaran yang lebih bersifat personal yang pada akhirnya dapat meningkatkan hasil belajar peserta didik.

Pemerolehan Tingkat pengetahuan dan keterampilan peserta didik secara umum dilakukan dengan ujian. Ujian dilakukan secara berkala untuk dapat mengetahui proses pembelajaran dan perkembangan peserta didik. Beberapa manfaat yang didapat dengan pelaksanaan ujian adalah: 1) guru dapat mengukur pencapaian kompetensi siswa, 2) sebagai penentu kelulusan dan melakukan pemetaan Pendidikan, dan 3) menyediakan umpan balik bagi pihak pemangku kepentingan sekolah. Selain itu, ujian juga dapat memberikan manfaat bagi siswa, diantaranya adalah: 1) mengajarkan disiplin dan pengelolaan waktu, 2) membangun karakter dan kejujuran, dan 3) meningkatkan budaya belajar. Tetapi semua itu hanya dapat dicapai apabila kualitas ujian juga baik.

Dalam praktiknya, ujian tidak hanya ditentukan oleh soal yang diberikan saja, tetapi juga terpengaruh oleh beberapa hal. Ujian yang tidak diperhatikan dan disusun dengan baik akan menghasilkan sebuah bentuk tes yang akurasi pengukurannya rendah, dan akibatnya, tidak menjamin kualitas lulusannya. Sebaliknya, soal dapat dikatakan baik apabila memenuhi beberapa kriteria, seperti validitas, reliabilitas, Tingkat kesukaran, daya pembeda, praktikabilitas, dan kesesuaiannya dengan kurikulum. Soal yang baik akan memberikan hasil yang baik bagi semua pihak, baik peserta didik, guru, maupun pihak institusi pendidikan.

Hasil ujian juga dapat dijadikan sebuah database yang dapat digunakan untuk meningkatkan kualitas ujian. Dengan metode *Data Driven Decision Making* (DDDM), *database* dapat diolah sedemikian rupa sehingga hasilnya dapat digunakan untuk tujuan yang beragam, tidak hanya bagi institusi secara internal, tetapi juga secara eksternal [5]. Proses dalam metode DDDM dapat diukur dengan berbagai cara, salah satunya adalah dengan menggunakan algoritma Fuzzy.

Algoritma Fuzzy akan membantu dalam perhitungan yang tidak melulu benar atau salah saja. Logika fuzzy menjembatani bahasa mesin yang presisi dengan bahasa manusia yang menekankan pada makna atau arti (*significance*) [6]. Algoritma ini digunakan dalam menghitung validitas, reliabilitas, penentuan Tingkat (level) kesukaran dan Daya pembeda sebuah set soal ujian [7] [8].

Metode ini dipilih karena memberikan penilaian dengan beberapa elemen pembandingan dan memiliki fungsi fuzzifikasi.

Penelitian ini juga menyasar ujian bahasa Inggris yang memiliki 5 kategori (*listening, reading, grammar, social skills, dan picture/audio*) sehingga metode fuzzy dapat diterapkan pada tiap-tiap kategori dan secara keseluruhan (*overall*). Hal ini menjadikan penelitian ini menjadi lebih spesifik dan mendetail jika dibandingkan dengan penelitian-penelitian sejenis.

Selanjutnya, penelitian ini juga akan diarahkan pada penentuan kata kunci soal (*question's keyword*) yang nantinya dapat digunakan dalam sistem otomatisasi pemanggilan soal berdasarkan levelnya pada ujian berbasis online. Sistem ini akan menggunakan arahan *Computer Adaptive Testing* (CAT) sehingga soal akan diberikan secara dinamis dan berbeda pada tiap-tiap pesertanya.

Data Driven Decision Making (DDDM) telah digunakan dalam dunia Pendidikan, terdapat beberapa penelitian yang telah dilakukan terkait hal tersebut, diantaranya adalah dari [9] yang meneliti DDDM dari sudut pandang validitas dan reliabilitasnya. Hasil penelitian pada 179 administrator sekolah di Turkey menunjukkan bahwa DDDM adalah sebuah alat pengukuran yang mampu menghasilkan pengukuran yang valid dan terpercaya dan dapat digunakan dalam pengukuran berbagai keperluan di sekolah. Selanjutnya, [10] menunjukkan bahwa sekolah-sekolah telah banyak menggunakan *data analytics tools* termasuk untuk data mining, analytics cara belajar, dan *business intelligent* untuk mendapatkan masukan-masukan, rekomendasi, dan pengetahuan dari data Pendidikan. DDDM memberikan penegasan pada teori-teori pedagogi, framework, dan berbagai fenomena Pendidikan. Selanjutnya, Kaspi dan Venkatraman [11] yang meneliti tentang perhitungan pada hasil asesmen di perguruan tinggi pada mahasiswa Teknologi Informasi di Australia. Hasil perbandingan menyebutkan bahwa DDDM mampu memberikan hasil yang lebih akurat sehingga dapat digunakan sebagai dasar pengembangan metode Pendidikan di transformasi Pendidikan saat ini.

Di sisi lain, hasil dari perhitungan dengan metode DDDM atau Data-Based Decision Making (DBDM) tidak selamanya baik, hal ini bergantung pada karakteristik yang dimiliki oleh institusi pendidikan masing-masing. Implementasi hasil dapat berpengaruh pada kesuksesan Pendidikan jika terdapat kepemimpinan yang kuat, transparansi, standarisasi proses kerja, keterlibatan staf secara terus-menerus, dan pemandu akademik yang ulet [12]. Hal ini juga didukung oleh [13] yang menyebutkan bahwa DDDM hanya akan berguna jika system, para pengguna dan operator, juga para pemangku Keputusan dapat memahami secara menyeluruh semua hasil yang diberikan dari perhitungan data yang ada.

Standar soal ujian yang baik juga menjadi perhatian para peneliti bidang pendidikan. Integritas dari sebuah ujian yang bergantung pada standar kualitas ujian yang diberikan. Item analysis yang dilakukan pada sebuah ujian terdiri dari Tingkat kesukaran soal, validitas dan reliabilitasnya. Hasil penelitian menyebutkan bahwa standar ujian dapat mengurangi diskriminasi terhadap hasil kemampuan peserta [14] [15]. Hal senada juga diungkapkan oleh [16] dan [17] yang meneliti tentang fungsi Tingkat kesukaran, daya pembeda dan efisiensi distraktor pada soal ujian pilihan berganda. Hasilnya menyebutkan bahwa soal ujian yang kurang baik akan memberikan hasil yang kurang baik pula pada peserta.

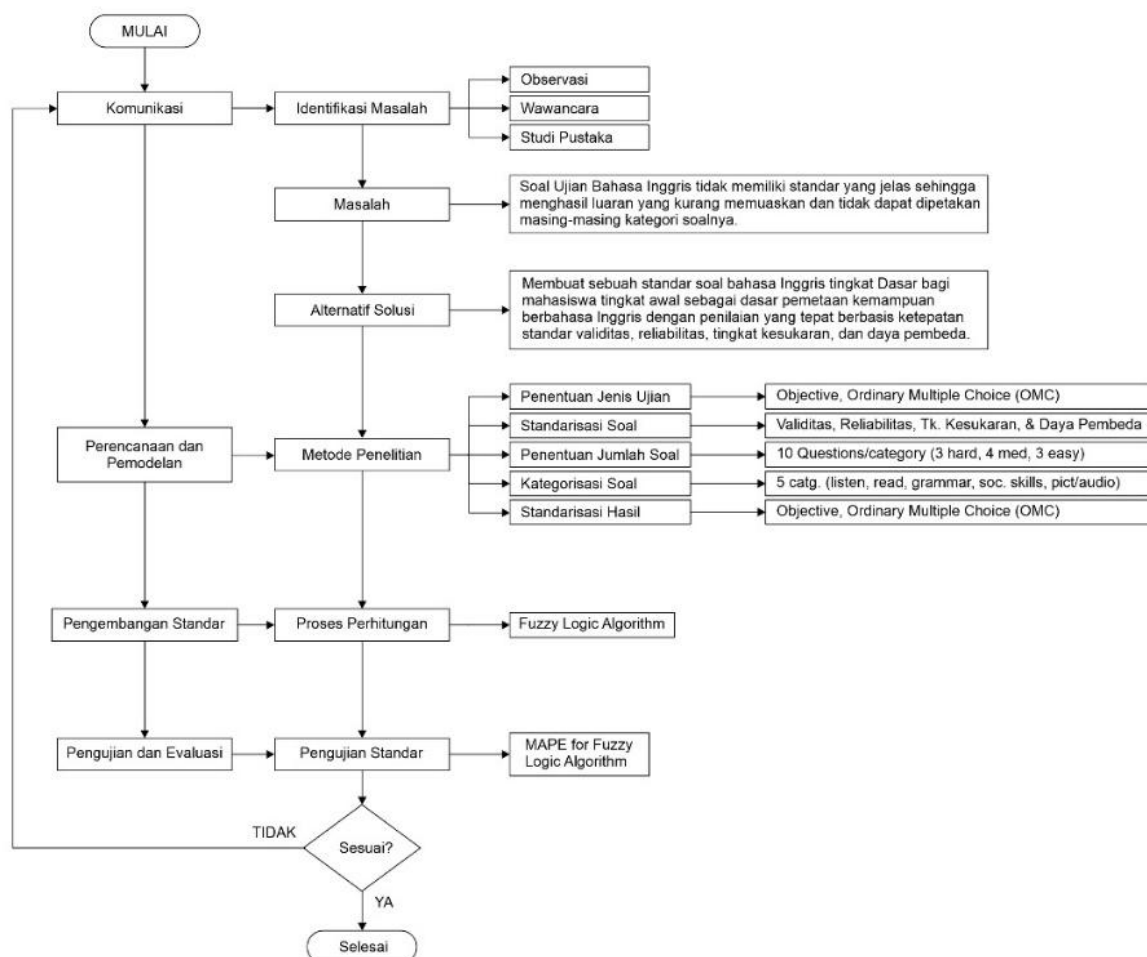
Keunikan algoritma fuzzy membuatnya dipakai untuk berbagai macam keperluan perhitungan. [18] menggunakan algoritma ini untuk menentukan kelayakan penerima beasiswa mahasiswa di sebuah universitas. Algoritma Fuzzy juga diterapkan pada beberapa penelitian terkait standar soal ujian. Pengukuran kompetensi siswa menggunakan Fuzzy dilakukan oleh [19] yang memanfaatkan perhitungan fuzzy untuk penelusuran butir soal dan pengukuran kemampuan peserta ujian. Hal tersebut juga pernah dilakukan oleh [20] yang menggunakan fuzzy untuk mengukur Tingkat

kemampuan pemecahan masalah matematis untuk siswa SMP. Selain itu, algoritma ini juga digunakan untuk menganalisis validitas, reliabilitas, Tingkat kesulitan, dan daya beda butir soal pada soal ujian akhir semester Tingkat sekolah dasar [21].

Dari semua penelitian di atas, penelitian ini memberikan sebuah khazanah baru pada pengukuran standar soal ujian Bahasa Inggris untuk perguruan tinggi dengan menggunakan Data Driven Decision Making (DDDM) dengan penerapan Algoritma Fuzzy untuk mengukur tingkat kesukaran, daya beda, validitas, dan reliabilitas soal, baik secara individu maupun bersama-sama.

2. METODOLOGI

Penelitian ini dilakukan di sebuah perguruan tinggi swasta yang tidak memiliki fakultas Bahasa sehingga model pembelajaran Bahasa Inggrisnya menggunakan *English for Specific Purposes* (ESP). Alur penelitian dimulai dengan identifikasi permasalahan melalui observasi dan wawancara yang kemudian didukung oleh data ujian terdahulu. Hasil observasi dan wawancara kemudian diolah untuk dijadikan standar awal penentuan validitas dan reliabilitas soal. Data hasil ujian terdahulu diolah untuk mendapatkan Tingkat kesukaran soal dan daya pembedanya. Peneliti juga ingin mengetahui keterkaitan antara daya pembeda soal dan validitasnya dalam penentuan kelayakan soal.



Gambar 1. Alur Penelitian

Peneliti mengadakan ujian Bahasa Inggris dasar pada 20 orang mahasiswa semester awal dengan tujuan untuk mendapatkan hasil ujian yang murni dengan peserta yang belum pernah menempuh pembelajaran Bahasa Inggris Tingkat perguruan tinggi sebelumnya. Penelitian ini menggunakan metode kuantitatif dan Teknik dokumentasi. Data penelitian berupa data lembar soal dan lembar jawaban Ujian Bahasa Inggris. Data yang didapatkan dari ujian tersebut kemudian digunakan sebagai data awal dalam model perhitungan dengan menggunakan *Data Driven Decision Making* (DDDM).

Data tersebut kemudian akan digunakan sebagai awal perhitungan dan acuan dalam penentuan validitas, reliabilitas, Tingkat kesukaran soal, dan daya pembeda. Kesemua data tersebut akan dihitung menggunakan algoritma Fuzzy. Hasil yang didapat diharapkan dapat menjadi acuan dalam peningkatan kualitas soal-soal ujian beserta berbagai macam pengembangannya.

Standar soal ujian ditentukan dalam sebuah *focus group discussion* (FGD) yang diikuti oleh beberapa pemangku kepentingan, di antaranya adalah para dosen pengampu mata kuliah Bahasa Inggris, bagian akademik kampus, mahasiswa, dan ahli bahasa dari luar lingkungan kampus. Hasil yang didapatkan adalah bahwa sebuah soal ujian harus memiliki tingkat kesukaran (*difficulty level*), daya beda (*differentiation force*), validitas, dan reliabilitas yang terukur sehingga menghasilkan sebuah tes yang tidak hanya baik bagi pihak dosen, tetapi juga adil bagi pihak peserta ujian. Dalam praktiknya, soal-soal ujian ini harus dapat dihitung dan ditentukan dengan cepat. Sebagai solusinya, DDDM database dan algoritma Fuzzy diputuskan sebagai dasar penentu standar soal ujian Bahasa Inggris.

Pada desain penelitian diatas, penulis melakukan tahapan penelitian sebagai berikut:

1. Pengumpulan data, pada level ini penulis menghimpunkan materi yang dibutuhkan dalam pembuatan penelitian ini. Data berasal dari hasil ujian bahasa Inggris dengan 5 kategori ujian. Ujian ini memiliki skor yang telah ditentukan sebelumnya.
2. Dilakukan penyesuaian jumlah soal/kategori, level, dan skor/level ujian.
- 3.

Tabel 1. Pembagian Jumlah Soal Berdasarkan Tingkat Kesukaran

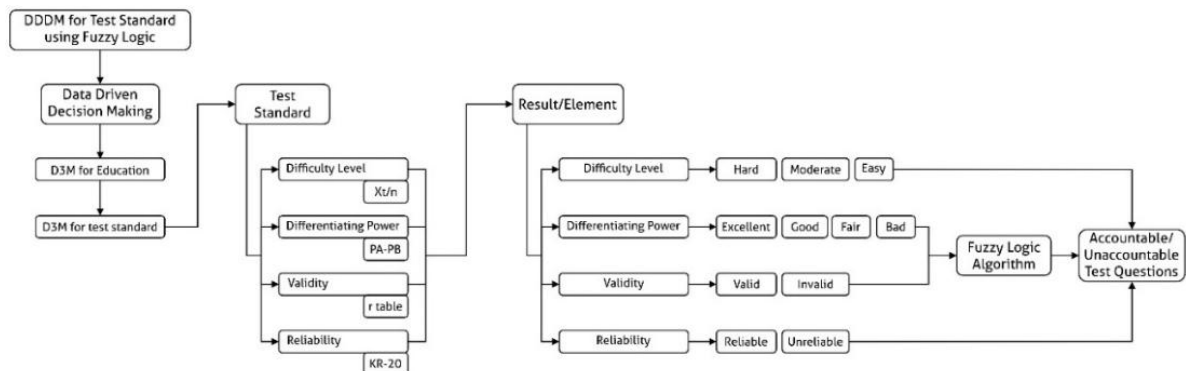
Soal	Easy	Moderate	Hard	Subtotal
Listening	3	4	3	10
Reading	3	4	3	10
Grammar	3	4	3	10
Social Skills	3	4	3	10
Picture/Audio	3	4	3	10
Total				50

Tabel 2. Standar Pembobotan Soal Ujian per Kategori

Soal	Jumlah soal	Poin/soal	Sub total
Easy	3	1,5	4,5
Moderate	4	2	8
Hard	3	2,5	7,5
Total			20

Dalam sistem ujian Bahasa Inggris yang dilaksanakan secara online, data yang dianalisis berupa data validitas, reliabilitas, tingkat kesukaran dan daya pembeda serta menggunakan rumus korelasi point biserial (validitas), rumus K-R 20 (reliabilitas), dan algoritma fuzzy logic. Bentuk tes

yang dianalisis adalah pilihan ganda. Validitas adalah indeks yang menunjukkan sejauh mana suatu alat penghitung itu menilai apa yang ingin diukur.



Gambar 2. Alur Pengukuran standar soal ujian menggunakan Fuzzy

Hasil dari pengukuran standar tersebut akan memberikan gambaran tentang soal-soal ujian yang diberikan dan kapabilitasnya dalam memenuhi standar soal yang diinginkan.

3. HASIL DAN PEMBAHASAN

Ujian dilaksanakan di sebuah kampus Teknik di Jakarta Barat. 20 orang mahasiswa Tingkat 1 melaksanakan ujian Bahasa Inggris dan hasilnya dipecah berdasarkan soal dan hasil jawaban para peserta. Hasil ujian dapat dilihat pada table 3 di bawah ini:

Tabel 3. Hasil Ujian Peserta

ST ID	QUESTION NUMBER																																																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	
ID1	1	1	0	1	0	0	1	0	1	0	1	1	0	1	1	0	1	0	0	0	0	0	0	1	0	1	0	0	1	0	1	1	0	1	1	0	1	0	1	1	1	1	1	0	1	0	0	0	0		
ID2	1	1	1	1	0	0	1	1	0	1	1	1	0	0	0	0	1	1	0	0	0	0	0	1	1	1	0	1	0	0	1	0	0	1	1	1	1	1	0	1	1	1	0	1	1	0	0	1	0		
ID3	1	1	1	1	1	1	0	1	1	1	0	1	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0	1	1	0		
ID4	1	1	0	1	0	1	1	1	1	0	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	0	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	0	
ID5	1	1	0	0	0	0	1	1	0	0	1	1	0	1	1	1	0	1	0	0	1	0	1	1	1	1	0	1	0	0	1	1	0	1	1	1	1	0	1	0	0	1	1	0	1	0	0	0	0		
ID6	1	1	0	1	0	1	1	1	1	0	1	1	1	1	0	0	1	1	1	0	1	1	0	1	1	1	1	0	0	1	1	0	0	1	1	1	1	0	0	1	0	0	1	1	1	1	1	0	1		
ID7	1	1	0	1	0	0	1	1	1	0	1	0	1	1	1	0	1	0	1	1	1	1	0	1	1	1	0	0	1	0	1	1	0	1	1	1	1	1	0	0	1	1	1	1	1	1	0	1	1		
ID8	1	1	1	1	1	1	0	1	0	1	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	1	1	0		
ID9	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0	0	1	1	1	0	0	1	1	1	0	1	0	0	0	0	0			
ID10	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1		
ID11	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
ID12	1	1	1	0	0	0	0	1	1	1	1	0	0	0	1	0	0	1	0	1	0	0	0	0	1	1	1	0	0	1	0	1	1	1	1	1	1	1	1	0	0	0	1	1	1	0	1	1	0	1	
ID13	1	1	1	0	0	0	0	1	1	0	0	1	1	0	0	0	0	1	1	1	0	0	1	1	1	0	0	0	0	0	1	1	0	1	1	1	1	0	0	1	0	1	1	1	0	1	0	1	0		
ID14	1	1	1	0	1	0	0	1	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0	1	0	1	0		
ID15	0	0	1	1	0	0	0	0	1	1	1	1	1	0	0	0	1	1	0	1	0	1	0	1	1	1	0	0	1	0	1	1	1	1	1	1	1	0	0	0	0	1	0	0	0	1	0	0	1	0	
ID16	1	1	1	1	0	0	1	0	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	0	0	0	1	1	1	1	1	1	0	1	0	1	0	
ID17	1	1	1	0	0	0	1	0	1	1	1	1	1	0	0	0	0	1	1	1	0	0	0	0	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	0	0	1	1	0	0	0	1	1	0		
ID18	0	1	1	1	1	1	1	1	0	1	1	0	1	1	0	0	0	1	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	
ID19	0	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
ID20	1	0	1	1	1	0	0	1	1	0	1	1	1	0	0	0	0	0	0	1	1	0	0	0	1	1	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
TOTAL	17	17	13	13	6	14	14	16	6	18	18	13	6	6	4	12	17	13	11	6	8	6	19	18	18	5	8	9	8	20	16	6	20	20	14	14	6	14	6	14	20	19	13	19	6	13	6	14	6		

3.1. Tingkat Kesukaran Soal (Level Difficulty)

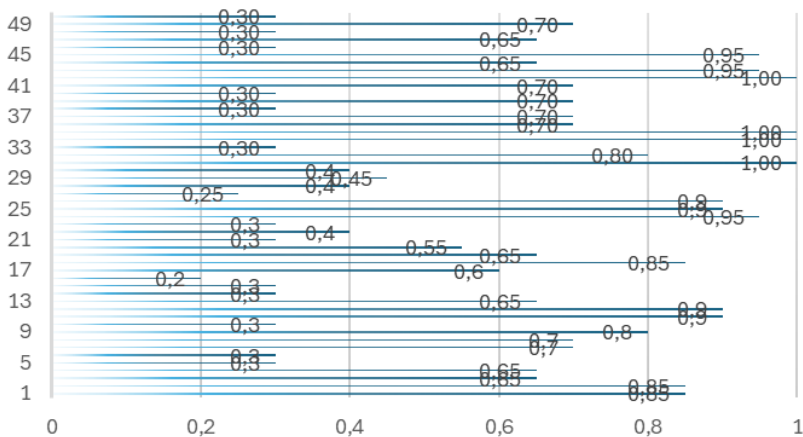
Hasil ujian tersebut kemudian dihitung satu persatu menggunakan rumus-rumus yang telah ditentukan sebelumnya. Rangkuman soal benar (*correct number summary*) kemudian dihitung menggunakan rumus $P = N_p/N$, dimana P adalah persentasi Tingkat kesukaran, N adalah jumlah soal benar, dan N adalah jumlah soal. Hasilnya dapat dilihat pada table 4 dan 5 di bawah ini:

Tabel 4. Rangkuman Tingkat Kesukaran Soal

Kategori	Nomor Soal										Sub-total
	1	2	3	4	5	6	7	8	9	10	
Listening	0,85	0,85	0,65	0,65	0,3	0,3	0,7	0,7	0,8	0,3	6,1
Reading	0,9	0,9	0,65	0,3	0,3	0,2	0,6	0,85	0,65	0,55	5,9
Grammar	0,3	0,4	0,3	0,95	0,9	0,9	0,25	0,4	0,45	0,4	5,25
Social Skills	1	0,8	0,3	1	1	0,7	0,7	0,3	0,7	0,3	6,8
Picture/Audio	0,7	1	0,95	0,65	0,95	0,3	0,65	0,3	0,7	0,3	6,5
											30,55

Tabel 5. Persentasi Tingkat Kesukaran Soal

Jumlah Soal	Kategori Soal		
50	Easy	Moderate	Difficult
Presentase	32%	38%	30%



Gambar 3. Grafik Tingkat Kesukaran Soal

Hasil yang didapat adalah:
Tingkat Kesukaran Perangkat Tes:

$$\left(Pp = \frac{\sum n.p}{n} \right) \tag{1}$$

maka,
 $Pp = 30.95/50 = 0,611$ – termasuk ke dalam kategori **Sedang/Moderate**.
Tipe soal dianggap sudah cukup IDEAL karena memiliki kategori SEDANG yang diharapkan mampu mengakomodir semua peserta (Kelompok ATAS & BAWAH).

3.2. Daya Beda (Differentiation Force)

Daya beda soal dibuat dengan menentukan terlebih dahulu batas atas dan batas bawah dari hasil ujian siswa. Hasil dari perhitungan dapat dilihat pada tabel 6 di bawah ini:

Tabel 6. Daya Beda

No	STD ID	SCORE	REMARK	No	STD ID	SCORE	REMARK
1	ID10	44	Atas	11	ID02	28	Bawah
2	ID11	43	Atas	12	ID12	28	Bawah
3	ID03	40	Atas	13	ID14	28	Bawah
4	ID08	38	Atas	14	ID20	28	Bawah
5	ID18	37	Atas	15	ID01	25	Bawah
6	ID04	35	Atas	16	ID05	25	Bawah
7	ID06	33	Atas	17	ID13	25	Bawah
8	ID07	33	Atas	18	ID15	24	Bawah
9	ID16	33	Atas	19	ID19	17	Bawah
10	ID17	31	Atas	20	ID09	16	Bawah

Interpretasi Daya beda dengan aturan seperti di bawah ini:

Tabel 7. Hasil Interpretasi Daya Beda

SOAL	DP	RESULT	SOAL	DP	RESULT	SOAL	DP	RESULT	SOAL	DP	RESULT	SOAL	DP	RESULT
1	0,10	Kurang Baik	11	0,90	Baik Sekali	21	0,60	Baik	31	0,00	Kurang Baik	41	0,00	Kurang Baik
2	0,30	Cukup	12	0,90	Baik Sekali	22	0,60	Baik	32	0,20	Cukup	42	0,00	Kurang Baik
3	0,10	Kurang Baik	13	0,65	Baik	23	0,20	Cukup	33	0,20	Cukup	43	0,10	Kurang Baik
4	0,50	Baik	14	0,30	Cukup	24	0,10	Kurang Baik	34	0,00	Kurang Baik	44	0,40	Baik
5	0,20	Cukup	15	0,30	Cukup	25	0,20	Cukup	35	0,00	Kurang Baik	45	0,10	Kurang Baik
6	0,60	Baik	16	0,20	Cukup	26	0,20	Cukup	36	0,40	Baik	46	0,20	Cukup
7	0,60	Baik	17	0,60	Baik	27	0,50	Baik	37	0,00	Kurang Baik	47	0,00	Kurang Baik
8	0,00	Kurang Baik	18	0,85	Baik Sekali	28	0,40	Baik	38	0,20	Cukup	48	0,20	Cukup
9	0,40	Baik	19	0,65	Baik	29	0,10	Kurang Baik	39	0,20	Cukup	49	0,60	Baik
10	0,00	Kurang Baik	20	0,55	Baik	30	0,40	Baik	40	0,20	Cukup	50	0,40	Baik

Dengan persamaan Daya Beda $(SA-SB)/IA \times 100\%$ dimana SA adalah siswa dengan batas atas dan SB adalah siswa dengan batas bawah, dan IA adalah Skor ideal kelompok atas, maka dihasilkan:

$$DP = \frac{(SA-SB)}{IA} \cdot 100\% \quad (2)$$

$$DP = (372-247)/619 \times 100\%$$

$$DP = 125/619 \times 100\%$$

$$DP = 0,202 = 20,2\% - \text{Kategori } \mathbf{Cukup}.$$

3.3. Perhitungan Validitas dan Reliabilitas

Perhitungan validitas dilakukan dengan menghubungkan korelasi masing-masing skor pertanyaan ujian dengan skor total masing-masing pertanyaan tersebut. Soal-soal yang berkorelasi secara signifikan terhadap skor total mengindikasikan bahwa soal tersebut sudah tepat. Rumus yang digunakan adalah:

$$r_{xy} = \frac{N \sum XY - \sum X \sum Y}{\sqrt{(N \sum X^2 - (\sum X)^2)(N \sum Y^2 - (\sum Y)^2)}} \quad (3)$$

Dimana:

r_{xy} = koefisien korelasi antara variabel X dan Y

N = jumlah responden

$\sum X$ = jumlah skor butir soal

$\sum Y$ = jumlah skor total soal

ΣX^2 = jumlah skor kuadrat butir soal

ΣY^2 = jumlah skor total kuadrat butir soal

Sementara hasil perhitungan reliabilitas dihitung menggunakan rumus KR-20, yaitu:

$$KR - 20 = \left(\frac{k}{(k-1)} \right) \cdot \left(\frac{1 - \Sigma p_j q_j}{\sigma^2} \right) \quad (4)$$

Dimana:

K: total jumlah pertanyaan

p_j : jumlah peserta yang menjawab pertanyaan dengan benar

q_j : jumlah peserta yang menjawab pertanyaan dengan salah

σ^2 : Variasi skor peserta yang ikut ujian

Hasil yang didapat dari hasil perhitungan validitas dan reliabilitas dapat dilihat pada tabel 9 di bawah ini:

Tabel 8. Hasil Perhitungan Validitas dan Reliabilitas

SOAL	VALIDITAS	STATUS VALIDITAS	RELIABILITAS	SOAL	VALIDITAS	STATUS VALIDITAS	RELIABILITAS
1	0,31	Invalid	0,576	26	0,48	Invalid	0,581
2	0,45	Invalid	0,576	27	0,59	Valid	0,581
3	0,52	Invalid	0,576	28	0,54	Valid	0,581
4	0,63	Valid	0,576	29	0,20	Invalid	0,581
5	0,47	Valid	0,576	30	0,54	Valid	0,581
6	0,64	Valid	0,576	31	0,00	Invalid	0,417
7	0,43	Valid	0,576	32	0,26	Invalid	0,417
8	0,26	Invalid	0,576	33	0,43	Valid	0,417
9	0,60	Valid	0,576	34	0,00	Invalid	0,417
10	0,13	Valid	0,576	35	0,00	Invalid	0,417
11	0,35	Valid	0,506	36	0,37	Valid	0,417
12	0,17	Invalid	0,506	37	0,26	Invalid	0,417
13	0,43	Invalid	0,506	38	0,09	Invalid	0,417
14	0,46	Invalid	0,506	39	0,26	Invalid	0,417
15	0,10	Invalid	0,506	40	0,32	Invalid	0,417
16	0,66	Valid	0,506	41	0,15	Invalid	0,287
17	0,58	Invalid	0,506	42	0,00	Invalid	0,287
18	0,29	Valid	0,506	43	0,03	Invalid	0,287
19	0,55	Valid	0,506	44	0,26	Invalid	0,287
20	0,51	Invalid	0,506	45	0,15	Invalid	0,287
21	0,60	Valid	0,581	46	0,26	Invalid	0,287
22	0,65	Valid	0,581	47	-0,18	Invalid	0,287
23	0,20	Invalid	0,581	48	0,20	Invalid	0,287
24	0,15	Invalid	0,581	49	0,66	Valid	0,287
25	0,48	Valid	0,581	50	0,43	Valid	0,287

3.4. Perhitungan Fuzzy untuk Kelayakan Soal

Perhitungan kelayakan soal dibuat dengan membandingkan dua variabel standar soal yaitu daya beda (*differentiation force*) dan validitas soal ujian. Hal ini dimulai dengan menentukan fuzzy rules untuk kedua variabel tersebut untuk kemudian dihitung. Tabel fuzzy rules ditampilkan di bawah ini:

Tabel 9. Fuzzy Rules untuk Kelayakan Soal

R	Rules dengan Kata
R1	Jika DP Baik sekali dan Val Valid maka Soal LAYAK
R2	Jika DP Baik dan Val Valid maka Soal LAYAK
R3	Jika DP Cukup dan Val Valid maka Soal LAYAK
R4	Jika DP Kurang Baik dan Val Valid maka Soal TIDAK LAYAK

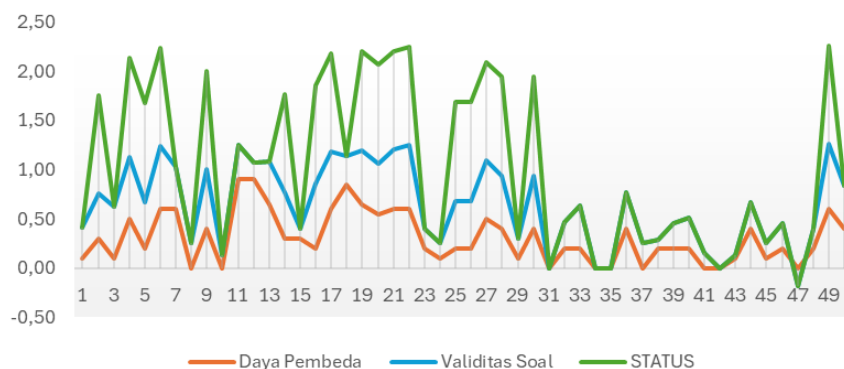
- R5 Jika DP Baik sekali dan Val Invalid maka Soal TIDAK LAYAK
- R6 Jika DP Baik dan Val Invalid maka Soal TIDAK LAYAK
- R7 Jika DP Cukup dan Val Invalid maka Soal TIDAK LAYAK
- R8 Jika DP Kurang Baik dan Val Invalid maka Soal TIDAK LAYAK

Dari *rules* tersebut didapatkan hasil sebagai berikut:

Tabel 10. Hasil perhitungan Fuzzy untuk Kelayakan Soal

SOAL	DP	RESULT	VAL	RESULT	KELAYAKAN	SOAL	DP	RESULT	VAL	RESULT	KELAYAKAN
1	0,10	Kurang Baik	0,31	Invalid	TIDAK LAYAK	26	0,20	Cukup	0,48	Valid	LAYAK
2	0,30	Cukup	0,45	Valid	LAYAK	27	0,50	Baik	0,59	Valid	LAYAK
3	0,10	Kurang Baik	0,52	Valid	TIDAK LAYAK	28	0,40	Baik	0,54	Valid	LAYAK
4	0,50	Baik	0,63	Valid	LAYAK	29	0,10	Kurang Baik	0,20	Invalid	TIDAK LAYAK
5	0,20	Cukup	0,47	Valid	LAYAK	30	0,40	Baik	0,54	Valid	LAYAK
6	0,60	Baik	0,64	Valid	LAYAK	31	0,00	Kurang Baik	0,00	Invalid	TIDAK LAYAK
7	0,60	Baik	0,43	Invalid	TIDAK LAYAK	32	0,20	Cukup	0,26	Invalid	TIDAK LAYAK
8	0,00	Kurang Baik	0,26	Invalid	TIDAK LAYAK	33	0,20	Cukup	0,43	Invalid	TIDAK LAYAK
9	0,40	Baik	0,60	Valid	LAYAK	34	0,00	Kurang Baik	0,00	Invalid	TIDAK LAYAK
10	0,00	Kurang Baik	0,13	Invalid	TIDAK LAYAK	35	0,00	Kurang Baik	0,00	Invalid	TIDAK LAYAK
11	0,90	Baik Sekali	0,35	Invalid	TIDAK LAYAK	36	0,40	Baik	0,37	Invalid	TIDAK LAYAK
12	0,90	Baik Sekali	0,17	Invalid	TIDAK LAYAK	37	0,00	Kurang Baik	0,26	Invalid	TIDAK LAYAK
13	0,65	Baik	0,43	Invalid	TIDAK LAYAK	38	0,20	Cukup	0,09	Invalid	TIDAK LAYAK
14	0,30	Cukup	0,46	Valid	LAYAK	39	0,20	Cukup	0,26	Invalid	TIDAK LAYAK
15	0,30	Cukup	0,10	Invalid	TIDAK LAYAK	40	0,20	Cukup	0,32	Invalid	TIDAK LAYAK
16	0,20	Cukup	0,66	Valid	LAYAK	41	0,00	Kurang Baik	0,15	Invalid	TIDAK LAYAK
17	0,60	Baik	0,58	Valid	LAYAK	42	0,00	Kurang Baik	0,00	Invalid	TIDAK LAYAK
18	0,85	Baik Sekali	0,29	Invalid	TIDAK LAYAK	43	0,10	Kurang Baik	0,03	Invalid	TIDAK LAYAK
19	0,65	Baik	0,55	Valid	LAYAK	44	0,40	Baik	0,26	Invalid	TIDAK LAYAK
20	0,55	Baik	0,51	Valid	LAYAK	45	0,10	Kurang Baik	0,15	Invalid	TIDAK LAYAK
21	0,60	Baik	0,60	Valid	LAYAK	46	0,20	Cukup	0,26	Invalid	TIDAK LAYAK
22	0,60	Baik	0,65	Valid	LAYAK	47	0,00	Kurang Baik	-0,18	Invalid	TIDAK LAYAK
23	0,20	Cukup	0,20	Invalid	TIDAK LAYAK	48	0,20	Cukup	0,20	Invalid	TIDAK LAYAK
24	0,10	Kurang Baik	0,15	Invalid	TIDAK LAYAK	49	0,60	Baik	0,66	Valid	LAYAK
25	0,20	Cukup	0,48	Valid	LAYAK	50	0,40	Baik	0,43	Invalid	TIDAK LAYAK

Dapat dilihat pada tabel 10, menghasilkan 18 soal dianggap layak sedangkan 32 soal dianggap kurang layak sehingga perlu dilakukan perbaikan. Agar lebih memperjelas hasil, diagram batang kelayakan butir soal dari perspektif Daya Beda dan Validitas ditampilkan pada gambar 3.



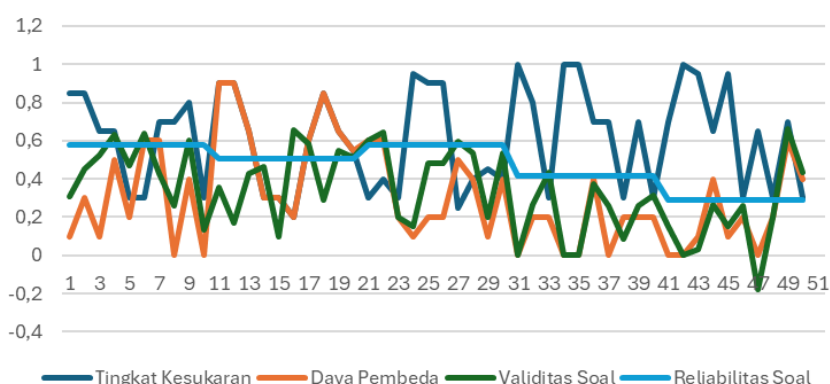
Gambar 4. Kelayakan Butir Soal menurut Status Kelayakan, Daya Beda, dan Validitas

Dapat dilihat pada gambar 4 di atas, bahwa jika daya beda rendah, dan validitas rendah, maka akan menghasilkan status yang rendah pula (tidak layak), sementara untuk status layak dapat dilihat jika soal dianggap valid dan memiliki setidaknya status daya beda yang cukup, baik, atau sangat baik.

Dari sudut pandang Tingkat kesukaran soal, dibandingkan dengan Daya beda, validitas, dan reliabilitas soal maka akan menghasilkan tabel seperti di bawah ini:

Tabel 11. Perbandingan Hasil Tingkat Kesukaran, Daya Beda, Validitas dan Reliabilitas

SOAL	Tingkat Kesukaran	Daya Pembeda	Validitas Soal	Reliabilitas Soal	SOAL	Tingkat Kesukaran	Daya Pembeda	Validitas Soal	Reliabilitas Soal	SOAL	Tingkat Kesukaran	Daya Pembeda	Validitas Soal	Reliabilitas Soal
1	0,85	0,10	0,31	0,576	21	0,3	0,60	0,60	0,581	41	0,70	0,00	0,15	0,287
2	0,85	0,30	0,45	0,576	22	0,4	0,60	0,65	0,581	42	1,00	0,00	0,00	0,287
3	0,65	0,10	0,52	0,576	23	0,3	0,20	0,20	0,581	43	0,95	0,10	0,03	0,287
4	0,65	0,50	0,63	0,576	24	0,95	0,10	0,15	0,581	44	0,65	0,40	0,26	0,287
5	0,3	0,20	0,47	0,576	25	0,9	0,20	0,48	0,581	45	0,95	0,10	0,15	0,287
6	0,3	0,60	0,64	0,576	26	0,9	0,20	0,48	0,581	46	0,30	0,20	0,26	0,287
7	0,7	0,60	0,43	0,576	27	0,25	0,50	0,59	0,581	47	0,65	0,00	-0,18	0,287
8	0,7	0,00	0,26	0,576	28	0,4	0,40	0,54	0,581	48	0,30	0,20	0,20	0,287
9	0,8	0,40	0,60	0,576	29	0,45	0,10	0,20	0,581	49	0,70	0,60	0,66	0,287
10	0,3	0,00	0,13	0,576	30	0,4	0,40	0,54	0,581	50	0,30	0,40	0,43	0,287
11	0,9	0,90	0,35	0,506	31	1,00	0,00	0,00	0,417					
12	0,9	0,90	0,17	0,506	32	0,80	0,20	0,26	0,417					
13	0,65	0,65	0,43	0,506	33	0,30	0,20	0,43	0,417					
14	0,3	0,30	0,46	0,506	34	1,00	0,00	0,00	0,417					
15	0,3	0,30	0,10	0,506	35	1,00	0,00	0,00	0,417					
16	0,2	0,20	0,66	0,506	36	0,70	0,40	0,37	0,417					
17	0,6	0,60	0,58	0,506	37	0,70	0,00	0,26	0,417					
18	0,85	0,85	0,29	0,506	38	0,30	0,20	0,09	0,417					
19	0,65	0,65	0,55	0,506	39	0,70	0,20	0,26	0,417					
20	0,55	0,55	0,51	0,506	40	0,30	0,20	0,32	0,417					



Gambar 5. Standar Soal dengan 4 Variabel Pengukuran

4. KESIMPULAN

Dari hasil perhitungan menggunakan algoritma Fuzzy untuk Tingkat kesukaran, didapatkan bahwa dari 50 soal yang diujikan, terdapat 16 soal mudah (32%), 19 soal sedang (38%), dan 15 soal sulit (30%) sehingga dalam perhitungan Tingkat kesukaran soal, tipe soal ini masuk dalam kategori sedang. Dalam perhitungan Daya beda didapatkan 3 soal baik sekali (6%), 17 soal baik (34%), 15 soal cukup (30%), dan 15 soal buruk (30%), dan secara general disebut memiliki tingkat kesukaran 0,611 (sedang). Dalam perhitungan persamaan Daya Beda menghasilkan 20,2%, termasuk dalam kategori cukup. Perhitungan Validitas menghasilkan 19 soal valid dan 31 soal tidak valid. Dalam perhitungan Reliabilitas didapatkan hasil 0,87 dan dianggap reliable. Penulis juga menghitung kelayakan soal yang didapatkan dari perhitungan fuzzy antara daya beda dan validitas soal, Dimana menghasilkan 18 soal layak (36%) dan 32 soal tidak layak (64%). Hal ini menunjukkan bahwa dalam kasus ini, terdapat 32 soal yang perlu diubah sehingga dapat membuat soal lebih layak untuk diujikan. Kesimpulan dari penelitian ini adalah bahwa metode Data Driven Decision Making (DDDM) dapat digunakan untuk mengkitung Tingkat kelayakan soal sehingga dapat memberikan masukan-masukan pada pengembangan standarisasi soal ujian, khususnya untuk bidang Bahasa Inggris. Penelitian selanjutnya sebaiknya dilakukan dengan data yang lebih besar dan pada beberapa lokasi dengan

tingkat pengetahuan peserta yang berbeda-beda, sehingga standar kelayakan soal, khususnya bagian validitas dan reliabilitas juga dapat lebih akurat. Penggunaan algoritma atau model perhitungan yang berbeda juga disarankan sehingga dapat memberikan manfaat pada pengembangan bidang ini.

UCAPAN TERIMA KASIH

Tim penulis mengucapkan terima kasih kepada Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Dirjen DIKTI, Kemdikbudristek) Republik Indonesia, Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) Institut Teknologi PLN (ITPLN) yang telah memberikan dukungan baik moril maupun materiil kepada tim penulis.

DAFTAR PUSTAKA

- [1] E. B. Mandinach and K. Schildkamp, "Misconceptions about data-based decision making in education: An exploration of the literature," *Stud. Educ. Eval.*, vol. 69, no. September 2019, p. 100842, 2021, doi: 10.1016/j.stueduc.2020.100842.
- [2] Y. Zhu, "A data driven educational decision support system," *Int. J. Emerg. Technol. Learn.*, vol. 13, no. 11, pp. 4–16, 2018, doi: 10.3991/ijet.v13i11.9582.
- [3] M. Nurzen S, "Pengambilan keputusan berbasis data untuk perencanaan pendidikan: strategi untuk keberhasilan kepala sekolah," *J. Konseling dan Pendidik.*, vol. 10, no. 4, pp. 589–596, 2022, doi: 10.29210/189500.
- [4] S. Hong, Carrie Eunyoung Lawrence, "Action Research in Teacher Education: Classroom Inquiry, Reflection, and Data-Driven Decision Making.," *J. Inq. Action Educ.*, vol. 4, no. 2, pp. 1–17, 2011.
- [5] F. M. Van der Kleij, J. A. Vermeulen, K. Schildkamp, and T. J. H. M. Eggen, "Integrating data-based decision making, Assessment for Learning and diagnostic testing in formative assessment," *Assess. Educ. Princ. Policy Pract.*, vol. 22, no. 3, pp. 324–343, 2015, doi: 10.1080/0969594X.2014.999024.
- [6] A. Mreła, O. Sokolov, and W. Urbaniak, "The method of learning outcomes assessment based on fuzzy relations," *Bull. Polish Acad. Sci. Tech. Sci.*, vol. 67, no. 3, pp. 527–533, 2019, doi: 10.24425/bpasts.2019.129651.
- [7] D. H. Rao, S. R. Mangalwede, and V. B. Deshmukh, "Student performance evaluation model based on scoring rubric tool for network analysis subject using fuzzy logic," *Int. Conf. Electr. Electron. Commun. Comput. Technol. Optim. Tech. ICEECCOT 2017*, vol. 2018-January, no. May, pp. 857–861, 2017, doi: 10.1109/ICEECCOT.2017.8284623.
- [8] J. Domingo-Peña, H. Martínez-García, and E. García-Vílchez, "a Contribution To Students' Assessment Adjusts of Multiple Choice Questionnaires With Fuzzy Logic," *SEFI 2022 - 50th Annu. Conf. Eur. Soc. Eng. Educ. Proc.*, pp. 520–530, 2022, doi: 10.5821/conference-9788412322262.1377.
- [9] E. Doğan and A. O. Demirbolat, "Data-driven Decision-Making in Schools Scale: A Study of Validity and Reliability conditions of the Creative Commons Attribution license (CC BY-NC-ND)," *Int. J. Curric. Instr.*, vol. 13, no. 1, pp. 507–523, 2021.
- [10] S. Gaftandzhieva, S. Hussain, S. Hilčenko, R. Doneva, and K. Boykova, "Data-driven Decision Making in Higher Education Institutions: State-of-play," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 6, pp. 397–405, 2023, doi: 10.14569/IJACSA.2023.0140642.
- [11] S. Kaspi and S. Venkatraman, "Data-Driven Decision-Making (DDDM) for Higher Education Assessments: A Case Study," *Systems*, vol. 11, no. 6, 2023, doi: 10.3390/systems11060306.

- [12] M. van Geel, A. J. Visscher, and B. Teunis, "School characteristics influencing the implementation of a data-based decision making intervention," *Sch. Eff. Sch. Improv.*, vol. 28, no. 3, pp. 443–462, 2017, doi: 10.1080/09243453.2017.1314972.
- [13] K. Schildkamp, "Data-based decision-making for school improvement: Research insights and gaps," *Educ. Res.*, vol. 61, no. 3, pp. 257–273, 2019, doi: 10.1080/00131881.2019.1625716.
- [14] A. H. Elgadal and A. A. Mariod, "Item Analysis of Multiple-choice Questions (MCQs): Assessment Tool For Quality Assurance Measures," *Sudan J. Med. Sci.*, vol. 16, no. 3, pp. 334–346, 2021, doi: 10.18502/sjms.v16i3.9695.
- [15] D. Kumar, R. Jaipurkar, A. Shekhar, G. Sikri, and V. Srinivas, "Item analysis of multiple choice questions: A quality assurance test for an assessment tool," *Med. J. Armed Forces India*, vol. 77, pp. S85–S89, 2021, doi: 10.1016/j.mjafi.2020.11.007.
- [16] M. Marsevani, "Item Analysis of Multiple-Choice Questions: an Assessment of Young Learners," *English Rev. J. English Educ.*, vol. 10, no. 2, pp. 401–408, 2022, doi: 10.25134/erjee.v10i2.6241.
- [17] R. P. Patil, S. E. Bahekar, M. D. Kulkarni, and M. S. Baig, "Evaluation of validity and reliability of multiple-choice questions in second MBBS competency-based medical education-based pharmacology examination of medical institute of India," *Int. J. Res. Med. Sci.*, vol. 10, no. 12, p. 2878, 2022, doi: 10.18203/2320-6012.ijrms20223091.
- [18] K. Tarwati, "Metode Fuzzy Logic Untuk Penentuan Kelayakan Penerima Beasiswa Mahasiswa Di Universitas Muhammadiyah Sukabumi," *J. Sist. Informasi*, vol. 1, no. 2, p. 56, 2020.
- [19] F. Firdaus, W. Susanti, Y. Desnelita, I. Irwan, and G. Gustientiedina, "Penerapan Algoritma Fuzzy Berbasis Expert System untuk Mengukur Kompetensi Mahasiswa," *SEMASTER Semin. Nas. Teknol. Inf. Ilmu Komput.*, vol. 1, no. 1, pp. 266–274, 2020, [Online]. Available: <https://journal.unilak.ac.id/index.php/Semaster/article/view/6136/2679>
- [20] S. Loka, "Instrumentasi Kemampuan Pemecahan Masalah Matematis: Analisis Reliabilitas, Validitas, Tingkat Kesukaran Dan Daya Beda Butir Soal," *Gema Wiralodra*, vol. 10, no. 1, pp. 41–52, 2019.
- [21] I. Magdalena, S. N. Fauziah, S. N. Faziah, and F. S. Nupus, "Analisis Validitas, Reliabilitas, Tingkat Kesulitan Dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan," *BINTANG J. Pendidik. dan Sains*, vol. 3, no. 2, pp. 198–214, 2021, [Online]. Available: <https://ejournal.stitpn.ac.id/index.php/bintang>