

**KOMPARASI BENTUK TES PILIHAN GANDA DENGAN TES MENJODOHKAN (*MATCHING TEST*) DITINJAU DARI TINGKAT KESUKARAN PADA MATA PELAJARAN BAHASA ARAB DI MAN WAIMITAL KABUPATEN SERAM BAGIAN BARAT.**

**ST. JUMAEDA<sup>1</sup>**

Dosen FITK IAIN Ambon  
e-mail: [Edha\\_amel77@yahoo.com](mailto:Edha_amel77@yahoo.com)

**Abstrak**

Penelitian ini bertujuan untuk mengetahui perbandingan bentuk tes pilihan ganda dengan tes menjodohkan ditinjau dari tingkat kesukaran butir soal. Penelitian ini merupakan penelitian kuantitatif, artinya semua gejala diukur menggunakan besaran kuantitatif yang kemudian digeneralisasikan atau simpulan ditarik dari interpretasi terhadap angka-angka yang dihasilkan oleh perhitungan statistik maupun matematik. Instrumen yang digunakan dalam penelitian ini adalah Instrumen tes hasil belajar Bahasa Arab siswa MAN Waimital. Bentuk intrumennya berupa pilihan ganda dan menjodohkan (*matching test*), kemudian skor dari tiap-tiap butir tersebut dianalisis dengan menggunakan analisis butir dengan tujuan untuk menentukan tingkat kesukaran butir.

Populasi dalam penelitian ini adalah seluruh Siswa Kelas XI Madrasah Aliyah Negeri (MAN) Waimital tahun akademik 2015/2016. Dalam pengambilan sampel penelitian dilakukan dengan teknik acak sederhana (*simple random sampling*). Sampel penelitian sebanyak 200 siswa.

Berdasarkan hasil analisis data dengan menggunakan uji-t menunjukkan bahwa untuk  $\alpha = 0,05$  dengan  $dk = 30 - 1$  adalah 2,224 sedangkan nilai  $t_{hitung} = 0,224$ , sehingga dapat disimpulkan bahwa tidak terdapat perbedaan tingkat kesukaran butir antara tes bentuk pilihan ganda dengan tes menjodohkan (*matching test*) pada mata pelajaran Bahasa Arab pada siswa kelas XI di Madrasah Aliyah Negeri Waimital Kabupaten Seram Bagian Barat

**Kata Kunci:** *Tingkat kesukaran butir, Pilihan Ganda, Menjodohkan.*

**A. Pendahuluan**

Dalam upaya peningkatan mutu pendidikan kegiatan evaluasi mempunyai peran yang sangat penting. Sebab melalui evaluasi, pendidik akan mengetahui gambaran kemampuan para Siswa yang dievaluasi. Dalam evaluasi selalu mengandung proses.

---

<sup>1</sup> Dosen Fakultas Ilmu Tarbiyah dan Keguruan Institut Agama Islam Negeri Ambon

Proses evaluasi harus tepat terhadap tujuan yang biasanya dinyatakan dalam bahasa perilaku. Dikarenakan tidak semua perilaku dapat dinyatakan dengan alat evaluasi yang sama, maka evaluasi menjadi salah satu hal yang sangat sulit dan menantang, yang harus disadari oleh para guru. Menurut Undang-Undang Republik Indonesia Nomor 20 tahun 2003 Tentang Sistem Pendidikan Nasional Pasal 57 ayat (I); Evaluasi merupakan salah satu upaya dilakukan dalam rangka pengendalian mutu pendidikan secara Nasional sebagai bentuk akuntabilitas penyelenggaraan pendidikan kepada pihak-pihak yang berkepentingan, diantaranya terhadap peserta didik, lembaga dan program pendidikan.<sup>2</sup>

Adapun kegiatan evaluasi mempunyai hubungan dengan kegiatan pengukuran. Dan pengukuran adalah salah satu tahapan sangat penting dalam suatu proses penelitian ilmiah. Melalui pengukuran akan dihasilkan data penelitian, kemudian berdasarkan data penelitian tersebut dibuat penafsiran, kesimpulan dan implikasinya. Karena pencapaian perkembangan Siswa perlu diukur terlebih dahulu, baik posisi Siswa sebagai individu maupun dalam kelompok. Hal yang demikian perlu disadari oleh guru karena pada umumnya masing – masing Siswa mempunyai kemampuan yang bervariasi. Ada Siswa yang memiliki kemampuan cepat menangkap materi dan ada yang lambat. Dan guru dapat mengevaluasi pertumbuhan kemampuan Siswa tersebut dengan mengetahui apa yang mereka kerjakan dari awal sampai akhir. Pencapaian belajar ini dapat dievaluasi melalui kegiatan pengukuran (*measurement*).

Mengukur pencapaian hasil belajar dapat melibatkan pengukuran secara kuantitatif yang menghasilkan data kuantitatif misalnya tes dan sekor, dan juga dapat mengukur data kualitatif yang menghasilkan deskripsi tentang subyek atau obyek yang diukur, misalnya rendah, medium, dan tinggi. Jadi kegiatan mengukur atau sering disebut pengukuran tidak lain adalah bagian evaluasi yang memiliki tujuan untuk menghasilkan data baik secara kuantitatif, maupun kualitatif.<sup>3</sup>

Pengukuran menurut Popham adalah sebuah proses pemberian angka terhadap suatu individu menurut karakteristik tertentu.<sup>4</sup> Pengukuran tidak melibatkan

---

<sup>2</sup> Undang-undang No.20 tahun 2003 tentang Sistem Pendidikan Nasional, h. 1-2.

<sup>3</sup> Sukardi, *Evaluasi Pendidikan Prinsip dan Operasionalnya* (Jakarta: Bumi Aksara, 2009), h. 2-3.

<sup>4</sup> Norman E. Gronlund, *Measurement and Evaluation in Teaching* (New York: Machmillan Publishing Company, 1985), h. 5.

pertimbangan mengenai baik buruknya atau nilai dari tingkah laku yang sedang diukur. Seperti halnya tes, pengukuran pun tidak menentukan siapa yang lulus dan siapa yang tidak lulus.

Pengukuran hanya membuahkan data kuantitatif mengenai apa yang hendak diukur. Dan kegiatan pengukuran memerlukan sebuah alat ukur yang disebut dengan instrumen. Instrumen merupakan alat ukur yang digunakan untuk mengukur dalam rangka pengumpulan data. Dalam pendidikan, instrumen yang digunakan untuk mengumpulkan data berupa tes dan non tes.

Tes merupakan alat ukur pengumpulan data yang mendorong peserta memberikan penampilan maksimal. Instrumen non tes merupakan alat ukur yang mendorong peserta untuk memberikan penampilan tipikal, yaitu melaporkan keadaan dirinya dengan memberikan respon secara jujur sesuai dengan pikiran dan perasaannya.<sup>5</sup> Adapun dua syarat utama yang harus dipenuhi oleh suatu instrumen penilaian, yaitu validitas dan reliabilitas.

Hubungan antara reliabilitas dengan validitas dapat dilihat dalam tiga hal, yakni: pertama, reliabilitas merupakan limit dari validitas. Ini mengandung makna bahwa interpretasi dan keputusan yang dihasilkan kurang valid jika hasil pengukuran tidak konsisten (tidak reliabel), kedua, tingginya reliabilitas bukan jaminan validitas, reliabilitas yang tinggi tidak menjamin hasil interpretasi dan keputusan yang dihasilkan memberikan validitas yang tinggi. Hal itu disebabkan karena banyaknya ragam validitas yang berhubungan dengan reliabilitas, dan ketiga, reliabilitas merupakan syarat perlu untuk validitas, artinya derajat reliabilitas mempengaruhi validitas keputusan.<sup>6</sup>

Secara teoritis, siswa dalam suatu kelas merupakan populasi atau kelompok yang keadaannya heterogen. Dengan demikian maka apabila dikenai sebuah tes akan tercermin hasilnya dalam suatu kurva normal. Sebagian siswa berada di daerah sedang, sebagian kecil berada di sebelah kiri dan sebagian kecil lainnya berada di daerah kanan kurva. Dan apabila keadaan setelah hasil analisis tes ternyata tidak seperti yang diharapkan maka sudah barang tentu ada sesuatu (apa-apa) dengan tes tersebut.<sup>7</sup>

---

<sup>5</sup> Purwanto, *Evaluasi Hasil Belajar* (Yogyakarta: Pustaka Pelajar, 2009), h. 56.

<sup>6</sup> Anthony J. Nitko, *Educational Assessment of Student* (New Jersey: Prentice-Hall, Inc. A. Simon & Schuster Company, Englewoods Cliffs, 2001), h. 63.

<sup>7</sup> Suharsimi Arikuto, *Dasar-dasar Evaluasi Pendidikan* (Jakarta :Bumi Aksara, 1999), h. 204.

Salah satu cara untuk mengantisipasi keadaan yang tidak normal itu adalah dengan jalan melakukan penganalisaan terhadap tes hasil belajar yang dijadikan alat pengukur dalam rangka mengukur keberhasilan belajar dari para peserta tes tersebut. Di sini tester perlu mengadakan penelusuran atau pelacakan terhadap butir-butir soal yang merupakan bagian tak terpisahkan dari tes hasil belajar sebagai totalitas. Hal ini dilakukan dengan tujuan untuk mengetahui, apakah butir-butir item yang membangun tes hasil belajar sudah dapat menjalankan fungsinya sebagai alat ukur hasil belajar yang memadai atau belum. Adapun penganalisaan tersebut dapat dilakukan dengan mengukur derajat kesukaran butir, daya pembeda dan reliabilitas item tes.

Dalam pengukuran pembelajaran Bahasa Arab , Salah satu instrumen evaluasi adalah tes atau *ikhtibâr*. Tes bahasa dirancang dan disusun sesuai dengan tujuan, materi dan sasaran pembelajaran itu sendiri. Tes inilah yang banyak dilakukan oleh tenaga pengajar/dosen, karena memang berkaitan dengan tugas edukatifnya, yakni memberi evaluasi dan nilai terhadap pemerolehan dan hasil belajar peserta didik.<sup>8</sup>

Tes kebahasaan merupakan sejumlah prosedur dan instrumen yang didesain secara sistematis, digunakan oleh tenaga pengajar dalam mengamati dan mengetahui performa dan kompetensi salah satu keterampilan bahasa peserta didik atau keseluruhannya, sesuai dengan ukuran kuantitatif tertentu dengan maksud mencapai tujuan tertentu pula. Pengerjaan tes sangat tergantung pada petunjuk yang diberikan, misalnya: melingkari atau memberi tanda silang pada salah satu huruf di depan pilihan jawaban, mencoret jawaban yang salah, menerangkan, mengisi titik-titik, dan sebagainya.<sup>9</sup>

Dalam penyusunan tes sendiri terdapat beberapa prosedur yang antara lain, *Pertama*, butir-butir atau kalimat soal hendaknya hanya disesuaikan dengan tujuan khusus yang telah ditetapkan. *Kedua*, soal yang dibuat hendaknya sesuai dengan tingkat kemampuan peserta didik, terutama jika berbentuk tes pemerolehan.

*Ketiga*, penyusunan tes hendaknya disertai petunjuk yang jelas, baik mengenai cara dan tempat menjawabnya serta lamanya waktu yang disediakan. *Keempat*, redaksi atau rumusan masing-masing soal harus jelas, tidak bersayap dan multiinterpretasi,

---

<sup>8</sup>Yayat Hidayat, *Studi Prinsip Dasar Metode Pengajaran Bahasa Arab* , h. 4. Diakses tanggal 15 Januari 2016, <http://arabicforall.or.id/studi-prinsip-dasar-metode-pengajaran-bahasa-arab>.

<sup>9</sup> *Ibid.*

terukur, dan diskriminatif. Kelima, waktu yang diberikan untuk menjawab soal harus sebanding dengan tingkat kesulitan dan banyak soal. Keenam, scoring penilaian harus obyektif berdasarkan proporsi yang ditetapkan, bukan berdasarkan rekaan, dan jauh dari subyektivitas penilai.

Madrasah Aliyah Negeri Waimital merupakan sebuah lembaga pendidikan yang berada di wilayah Kabupaten Seram Bagian Barat. Sebagai sebuah lembaga pendidikan menengah dibidang pendidikan agama Islam tentunya Bahasa Arab merupakan salah satu mata pelajaran yang sangat diutamakan, sebagai dasar untuk mempelajari al-Qur'an dan Hadits sebagai pedoman hidup ummat islam dalam membentuk ahlak siswa dan menciptakan siswa yang berkarakter dan berahlak karimah sesuai dengan amanat undang-undang.

Adapun bentuk tes obyektif, dalam Bahasa Arab antara lain: pilihan ganda (*multiple choise*), pilihan benar-salah, menjodohkan, dan melengkapi isian dengan jawaban yang bersifat tertutup. Akan tetapi di MAN Waimital bentuk tes objektif yang terbiasa digunakan adalah tes bentuk pilihan ganda, mencari pasangan dan melengkapi isian.

Pada mata pelajaran Bahasa Arab khususnya, kedua bentuk tes obyektif yaitu bentuk pilihan ganda dan mencari pasangan pada dasarnya hampir sama, karena pada soal pilihan ganda memiliki alternatif jawabannya lebih sedikit akan memiliki pengecoh atau *distractor*, artinya dalam soal pilihan ganda hanya ada satu alternatif jawaban yang paling benar dan yang lain adalah pengecoh yang berfungsi untuk mengalihkan perhatian peserta tes.

Sedangkan pada bentuk tes menjodohkan memiliki alternatif jawaban yang lebih banyak sehingga kesalahan menjodohkan dalam satu pasangan akan mengakibatkan kesalahan dalam menjodohkan pasangan yang lain. Sehingga untuk mengetahui tingkat kesukaran. Selama ini item tes pilihan ganda merupakan jenis tes obyektif yang paling banyak digunakan oleh para guru. Karena beberapa penelitian membuktikan bahwa tes ini mampu mengukur pengetahuan yang luas dengan tingkat domain yang bervariasi dan item pilihan ganda memiliki semua persyaratan sebagai tes yang baik, yakni dilihat dari validitas, reliabilitas, dan daya pembeda antara siswa yang berhasil dengan siswa yang tidak berhasil.

Pada umumnya, analisis butir berkenaan dengan beberapa besaran, meliputi (a) ukuran responden uji coba yakni banyaknya responden yang perlu dilibatkan di dalam uji coba, (b) ukuran kelompok responden sekor tinggi dan kelompok responden sekor rendah, (c) ukuran taraf sukar butir, dan (d) ukuran daya beda butir.

Di samping kajian secara teoretik, ada sejumlah kriteria empirik yang digunakan untuk menentukan nilai bagi besaran itu gunakan menemukan butir yang baik atau yang tidak baik.

## **1. Evaluasi**

Mehrens dan Lehman dalam Purwanto juga berpendapat bahwa evaluasi adalah proses merencanakan, memperoleh dan menyediakan informasi yang sangat diperlukan untuk membuat alternatif keputusan.<sup>10</sup> Popham memberikan pendapat bahwa ada dua hal yang perlu diperhatikan dalam evaluasi, yaitu mengetahui validitas dan reliabilitas perangkat tes yang akan digunakan. Validitas menunjukkan apa yang hendak diukur, sedangkan reliabilitas berkenaan dengan sekor Siswa sebagai peserta tes yang memiliki sekor yang sama ketika dites pada waktu dan tempat yang berbeda.<sup>11</sup>

Adapun fungsi evaluasi dalam dunia pendidikan tidak dapat dilepaskan dari tujuan evaluasi itu sendiri. Di dalam batasan tentang evaluasi pendidikan yang telah dijelaskan sebelumnya, tersirat bahwa tujuan evaluasi pendidikan adalah untuk mendapat data pembuktian yang akan menunjukkan sampai sejauh mana tingkat kemampuan dan keberhasilan Siswa dalam pencapaian tujuan-tujuan kurikuler. Disamping itu juga dapat digunakan oleh guru-guru atau pengawas pendidikan untuk mengukur atau menilai sampai sejauh mana keefektifan pengalaman-pengalaman mengajar, kegiatan-kegiatan belajar, dan metode-metode mengajar yang digunakan. Dengan demikian, dapat dikatakan betapa pentingnya peranan dan fungsi evaluasi dalam proses belajar-mengajar.

Evaluasi adalah sebuah proses tertentu yang digunakan untuk menafsirkan perkembangan peserta didik baik positif ataupun negatif mengenai kurikulum tertentu sebagaimana yang telah ditetapkan, dengan tujuan untuk membuat sebuah keputusan.

---

<sup>10</sup> Mehrens and Lehman dalam Ngalim Purwanto, *Prinsip-prinsip dan Teknik Evaluasi Pengajaran* (Bandung: PT. Remaja Rosdakarya. 2008), h. 3.

<sup>11</sup> W. James Popham, *Modern Educational Measurement* (Los Angeles: University of California, 1981), h. 371.

Jadi setiap kegiatan evaluasi merupakan suatu proses yang direncanakan untuk memperoleh informasi atau data, yang kemudian data tersebut akan digunakan untuk membuat sebuah keputusan. Dan sudah barang tentu bahwa informasi atau data yang dikumpulkan harus sesuai dan mendukung tujuan dari evaluasi yang telah direncanakan tersebut.

## **2. Pengukuran**

Hopkins dan Stanley memberikan definisi bahwa pengukuran adalah mengkonstruksi, mengatministrasi dan mensekor prosedur penilaian.<sup>12</sup> Sementara menurut Hamalik pengukuran dimaksudkan untuk memperoleh informasi yang kuat dan akurat tentang sesuatu yang diukur. Informasi itu akan membantu pihak-pihak yang berkepentingan untuk membuat keputusan, yang pada gilirannya berguna untuk menentukan tindakan yang akan dilakukan.<sup>13</sup> Sedangkan menurut Livingston ef. Al., Pengukuran adalah sebuah cara pemberian angka terhadap object, atribut, atau perilaku. sedangkan tes merupakan alat untuk menurut aturan dengan cara memberi angka pada performen masing-masing individu.<sup>14</sup>

Berdasarkan beberapa pengertian evaluasi, penilaian dan pengukuran yang dikemukakan di atas, maka jelaslah sudah bahwa evaluasi, penilaian, dan pengukuran merupakan tiga konsep yang berbeda. Namun demikian, dalam praktik terutama dalam dunia pendidikan, ketiga konsep tersebut sering dipraktikkan dalam satu rangkaian kegiatan. Sebagai contoh, pelaksanaan evaluasi di sekolah maka, di dalamnya terintegrasi kegiatan pengukuran dan penilaian.

## **3. Tes Sebagai Instrumen Pengumpulan Data Hasil Belajar**

Testing merupakan suatu prosedur sistematis untuk mengamati atau mendeskripsikan satu atau lebih karakteristik seseorang dengan menggunakan skala numerik atau kategorik.<sup>15</sup> Purwanto memberikan pendapatnya bahwa tes merupakan instrumen alat ukur untuk pengumpulan data dimana dalam memberikan respon atas

---

<sup>12</sup> Kenneth D. Hopkins, Julian C. Stanley and R.R. Hopkins, *Education and Psychological Measurement* (Nedham Height, Massa Cussets: Allyn and Bacon, 1990), h. 225-244

<sup>13</sup> Oemar Hamalik, *Evaluasi Kurikulum* (Bandung: Remaja Rosdakarya, 1993), h. 2

<sup>14</sup> Ronald B. Livingston *et al.*, *Measurement and Assesment in Education* (Boston: Pearson Education, Inc., Upper Saddla River, New Jersey, 2009), h. 3

<sup>15</sup> Lee J. Cronbach, *Essential of Psychological Testing* (New York: Harper & Row, 1984), h. 26.

pertanyaan dalam instrumen, peserta didorong untuk memberikan penampilan maksimalnya.<sup>16</sup> Peserta tes diminta untuk mengeluarkan segenap kemampuan yang dimilikinya dalam memberikan respon atas pertanyaan dalam tes. Penampilan maksimum yang ditunjukkan memberikan kesimpulan mengenai kemampuan atau penguasaan yang dimilikinya.

Membuat pertanyaan tes (alat evaluasi) tidak mudah, sebab tes atau pertanyaan merupakan alat untuk melihat perubahan kemampuan dan tingkah laku siswa setelah ia menerima pengajaran dari guru di sekolah. Alat evaluasi yang salah akan menggambarkan kemampuan dan tingkah laku yang salah pula. Oleh karena itu teknik penyusunan alat evaluasi penting dipertimbangkan agar memperoleh hasil yang obyektif.<sup>17</sup>

Dari beberapa pendapat tersebut di atas dapat disimpulkan bahwa tes adalah alat yang digunakan untuk mengukur kemampuan peserta didik dari suatu pelajaran yang telah ditentukan dalam waktu dan tingkat tertentu atau dengan kata lain tes adalah alat/instrumen/cara yang dapat dipergunakan atau prosedur yang ditempuh untuk mengukur dan menilai kemampuan peserta didik atau *testee*, yang berbentuk pemberian tugas atau serangkaian tugas yang dikerjakan oleh *testee*, dimana nilai yang dihasilkan melambangkan tingkah laku atau prestasi *testee* yang dapat dibandingkan dengan nilai-nilai yang dicapai oleh *testee* lainnya atau nilai standar tertentu.

#### **4. Analisis Butir Soal**

Menurut Sudjana analisis butir soal adalah pengkajian pertanyaan tes agar diperoleh perangkat pertanyaan yang memiliki kualitas yang memadai.<sup>18</sup> Adapun analisis soal ini dilakukan untuk mengetahui berfungsi tidaknya sebuah soal. Analisis pada umumnya dilakukan melalui dua cara, yaitu; analisis kualitatif (*qualitatif control*) dan analisis kuantitatif (*quantitatif control*). Analisis kualitatif sering pula dinamakan sebagai validitas logis (*logical validity*) yang dilakukan sebelum soal digunakan untuk melihat berfungsi tidaknya sebuah soal. Analisis soal secara kuantitatif sering pula dinamakan sebagai validitas empiris (*empiric validity*) yang dilakukan untuk melihat

---

<sup>16</sup> Purwanto, *oh. cit.*, h. 63.

<sup>17</sup> Nana Sudjana, *Dasar-dasar Proses Belajar Mengajar* (Bandung: Sinar Baru Algesindo, 2005), h.115.

<sup>18</sup> Sudjana, *oh. cit.*, h. 135.



lebih berfungsi tidanya sebuah soal setelah soal itu diujicobakan kepada sampel yang representatif.

Adapun salah satu tujuan diadakanya analisis menurut Surapranata adalah untuk meningkatkan kualitas soal, yaitu apakah suatu soal, 1). *Dapat diterima*, karena telah didukung oleh data statistik yang memadai, 2). *Diperbaiki*, karena terbukti terdapat beberapa kelemahan atau bahkan, 3). *Tidak digunakan sama sekali*, karena terbukti secara empiris tidak berfungsi sama sekali.<sup>19</sup> Analisis butir adalah kegiatan menganalisis item-item tes dengan tujuan untuk mengetahui kualitas butir-butir soal yang telah dibuat.

## 5. Tingkat Kesukaran Butir

Surapranata, tingkat kesukaran dibedakan menjadi tiga kategori; Yaitu, soal yang memiliki  $p < 0.3$  disebut dengan kategori sukar, soal yang memiliki  $p > 0.7$  disebut kategori mudah, dan soal yang memiliki  $p$  antara  $0.3$  sampai dengan  $0.7$  disebut kategori sedang.<sup>20</sup> Ketiga kategori itu digambarkan pada tabel di bawah ini:

Sedangkan menurut Dali taraf sukar batir sering ditentukan melalui proporsi jalaban betul dan proporsi jalaban salah. Makin besar proporsi jalaban betul,  $p$ , makin tidak sukar pada batir itu. Sebaliknya makin besar proporsi jawaban salah,  $q$ , makin sukar batir tersebut.<sup>21</sup>

Tabel 2.2 : Kategori Tingkat Kesukaran

Nilai p	Kategori
$p < 0.3$	Sukar
$0.3 \leq p \leq 0.7$	Sedang
$p > 0.7$	Mudah

Salah satu asumsi dasar yang digunakan dalam pengukuran adalah adanya perbedaan individu secara sistematis pada konstruk atau isi yang diukur oleh soal Tes dapat menunjukkan perbedaan individu ini. Ketika tidak satupun peserta tes dapat menjawab benar suatu soal ( $p = 0$ ) atau semua Siswa dapat menjawab benar suatu soal hal ini tentunya kita tidak dapat mendeteksi adanya perbedaan peserta tes. Hal ini

<sup>19</sup> *Ibid.*, h. 11.

<sup>20</sup> Surapranata, *oh. cit.*, h. 21.

<sup>21</sup> Dali S. Naga, *Probabilitas dan Sekor Pada Hipotesis Statistika* (Jakarta: UPT Taruma Negara, 2008), h. 63.

menunjukkan bahwa analisis tingkat kesukaran dapat digunakan sebagai suatu indikator untuk menentukan adanya perbedaan kemampuan peserta tes. Tingkat kesukaran 0 maupun 1 tidak memberikan kontribusi apapun terhadap perbedaan kemampuan peserta tes. Oleh karena itu, soal tersebut cenderung untuk tidak digunakan.

Secara umum menurut teori klasik, tingkat kesukaran dapat dinyatakan melalui beberapa cara, diantaranya (1). *Proporsi menjawab benar*, (2). *Skala kesukaran linier*, (3). *Indeks davis*, dan (4). *Sakala bivariat*. Proporsi jawaban benar ( $p$ ), yaitu jumlah peserta tes yang menjawab benar pada butir soal yang dianalisis dibandingkan dengan jumlah peserta tes seluruhnya, dan ini merupakan model analisis tingkat kesukaran yang paling umum digunakan. Persamaan yang digunakan untuk menentukan tingkat kesukaran dengan proporsi menjawab benar adalah:

$$p = \frac{\sum x}{S_m N}$$

Dimana  $p$  : Proporsi menjawab benar atau tingkat kesukaran

$\sum x$  : Banyaknya peserta tes yang menjawab benar

$S_m$  : Sekor maksimum

$N$  : Jumlah peserta tes.<sup>22</sup>

Adapun menurut Sudjana indeks kesukaran butir dapat dicari dengan menggunakan rumus sebagai berikut:

$$I = \frac{B}{N}$$

Dimana  $I$  = Indeks kesukaran untuk setiap butir

$B$  = Banyaknya Siswa yang menjawab benar setiap butir

$N$  = Banyaknya Siswa yang memberikan jawaban pada soal yang dimaksudkan.<sup>23</sup>

Tingkat kesukaran merupakan jumlah proporsi yang menunjukkan soal dengan kategori sukar, sedang serta mudah. Idealnya, tingkat kesukaran soal perlu disesuaikan dengan kemampuan peserta tes sehingga diperoleh informasi yang dapat digunakan sebagai alat perbaikan atau peningkatan program pembelajaran. Sehingga dapat dikatakan bahwa yang digunakan untuk memperoleh kualitas soal yang baik disamping

<sup>22</sup> *Ibid.*, h. 12.

<sup>23</sup> Sudjana, *oh. cit.*, h. 137.

memenuhi validitas dan reliabilitas adalah adanya keseimbangan dari tingkat kesukaran tersebut. Keseimbangan yang dimaksudkan adalah adanya soal-soal yang termasuk mudah, sedang, dan sukar secara proporsional. Tingkat kesukaran soal dipandang dari kesanggupan atau kemampuan Siswa dalam menjawabnya, bukan dari sudut guru sebagai pembuat soal. Persoalan penting dalam melakukan analisis tingkat kesukaran soal adalah penentuan proporsi dan kriteria soal yang termasuk mudah, sedang, dan sukar.

## **B. Metode Penelitian**

Penelitian merupakan penelitian kuantitatif, artinya semua gejala diukur menggunakan besaran kuantitatif yang kemudian digeneralisasikan atau simpulan ditarik dari interpretasi terhadap angka-angka yang dihasilkan oleh perhitungan statistik maupun matematik. Ini berbeda dengan penelitian kuantitatif yang tidak mendasarkan simpulan pada proses analisis terhadap gejala-gejala yang dikualifikasikan.

Instrumen yang digunakan dalam penelitian ini adalah Instrumen tes hasil belajar Bahasa Arab siswa MAN Waimital. Bentuk instrumennya berupa pilihan ganda dan menjodohkan (*matching test*), kemudian skor dari tiap-tiap butir tersebut dianalisis dengan menggunakan analisis butir dengan tujuan untuk menentukan tingkat kesukaran butir.

Proses pengambilan data dalam penelitian ini adalah, dengan memberikan naskah yang berisi butir-butir tes dalam bentuk pilihan ganda dan dalam bentuk tes menjodohkan (*matching test*) kepada setiap responden, dan kegiatan tesnya hanya dilakukan sekali saja. Adapun kedua bentuk tes tersebut berasal dari kisi-kisi dan indikator yang sama.

Populasi dalam penelitian ini adalah seluruh siswa Kelas XI Madrasah Aliyah Negeri (MAN) Waimital tahun akademik 2015/2016. Dalam pengambilan sampel penelitian dilakukan dengan teknik acak sederhana (*simple random sampling*), sebanyak 200 siswa.

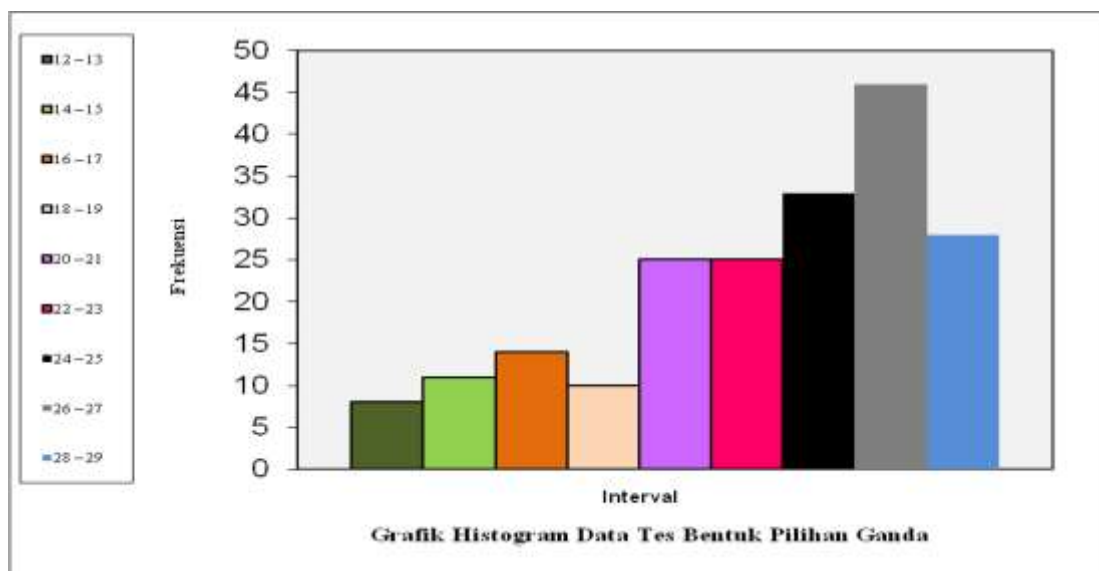
### C. Hasil Penelitian

Hasil perhitungan data penelitian mengenai sekor pilihan ganda pada mata pelajaran Bahasa Arab , memiliki rentang nilai 12 sampai 29, dengan nilai rata-rata (*mean*) sebesar 22,93; *modus* 26,338 dan *median* 23,924.

Deskripsi data sekor tes pilihan Siswa pada mata pelajaran Bahasa Arab ditunjukkan pada tabel dibawah ini:

**Tabel 4.1 Distribusi Frekuensi Sekor Bahasa Arab Pada Tes bentuk Pilihan Ganda**

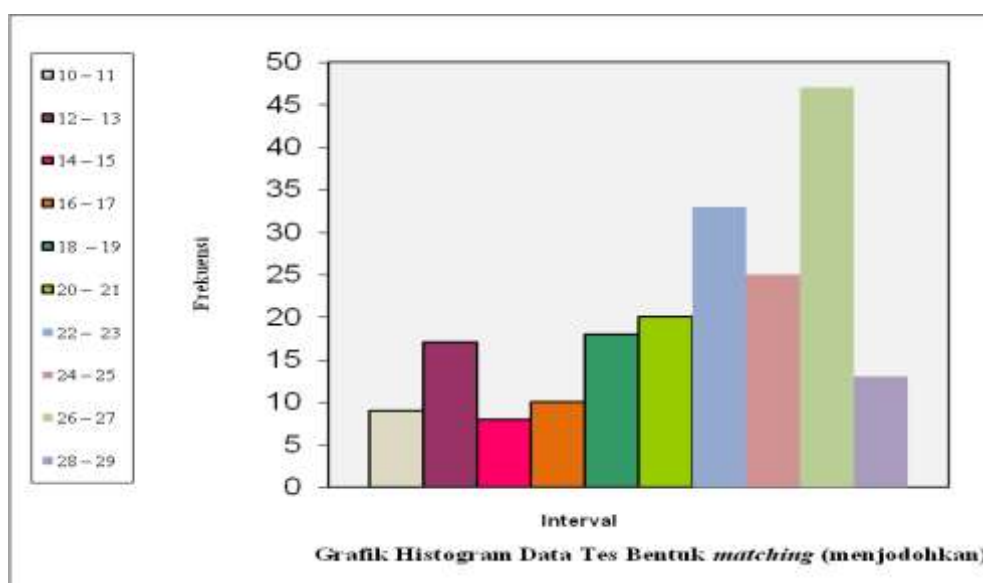
Interval	NilaiTengah	Batas Nyata	Frekuensi Absolut	Frekuensi Relatif
12 – 13	12,5	11,5 – 13,5	8	4%
14 – 15	14,5	13,5 – 15,5	11	5,5%
16 – 17	16,5	15,5 – 17,5	14	7%
18 – 19	18,5	17,5 – 19,5	10	5%
20 – 21	20,5	19,5 – 21,5	25	12,5%
22 – 23	22,5	21,5 – 23,5	25	12,5%
24 – 25	24,5	23,5 – 25,5	33	16,5%
26 – 27	26,5	25,5 – 27,5	46	23%
28 – 29	28,5	27,5 – 29,5	28	14%



Hasil perhitungan data penelitian mengenai sekor tes menjodohkan (*matching test*) pada mata pelajaran Bahasa Arab rentang nilai 10 sampai 28, dengan nilai rata-rata (*mean*) sebesar 21,47; *modus* 26,286 dan *median* 22,59.

**Tabel 4.2 Distribusi Frekuensi Sekor Bahasa Arab Pada Tes Bentuk Menjodohkan (*Matching test*)**

Interval	NilaiTengah	Batas Nyata	Frekuensi Absolut	Frekuensi Relatif
10 – 11	10,5	9,5 – 11,5	9	4,5%
12 – 13	12,5	11,5 – 13,5	17	8,5%
14 – 15	14,5	13,5 – 15,5	8	4%
16 – 17	16,5	15,5 – 17,5	10	5%
18 – 19	18,5	16,5 – 19,5	18	9%
20 – 21	20,5	19,5 – 21,5	20	10%
22 – 23	22,5	21,5 – 23,5	33	16,5%
24 – 25	24,5	23,5 – 25,5	25	12,5%
26 – 27	26,5	25,5 – 27,5	47	23,5%
28 – 29	28,5	27,5 -29,5	13	6,5%



Data tingkat kesukaran butir soal diperoleh berdasarkan hasil analisis dari data lembar jawaban tes pada masing-masing butir yang berbentuk pilihan ganda dan menjodohkan (*matching test*). Hal ini digunakan untuk mengetahui tingkat kesukaran masing-masing butir sebagai suatu indikator untuk menentukan adanya perbedaan kemampuan peserta tes sekor yang paling baik untuk membedakan tingkat kemampuan siswa antara siswa yang berkemampuan rendah dan siswa yang berkemampuan tinggi melalui analisis butir soal.

Berdasarkan pengelompokkan tingkat kesukaran butir soal antara tes bentuk pilihan ganda dengan menjodohkan (*matching*) sebagaimana tersebut di atas, maka dapat dikriteriakan sebagai berikut:

**Tabel 4.4: Tabel Kriteria tingkat kesukaran Butir Soal**

Nilai Tingkat Kesukaran	Kriteria soal
0.00 – 0.30	Sukar
0.30 – 0.70	Sedang
0.71 – 1.00	Mudah

Berdasarkan hasil penghitungan analisis butir soal pada jawaban siswa pada dua bentuk tes yaitu pilihan ganda dan tes menjodohkan (*matching test*), maka distribusi frekuensi pada masing-masing model penskoran dapat digambarkan pada tabel berikut :

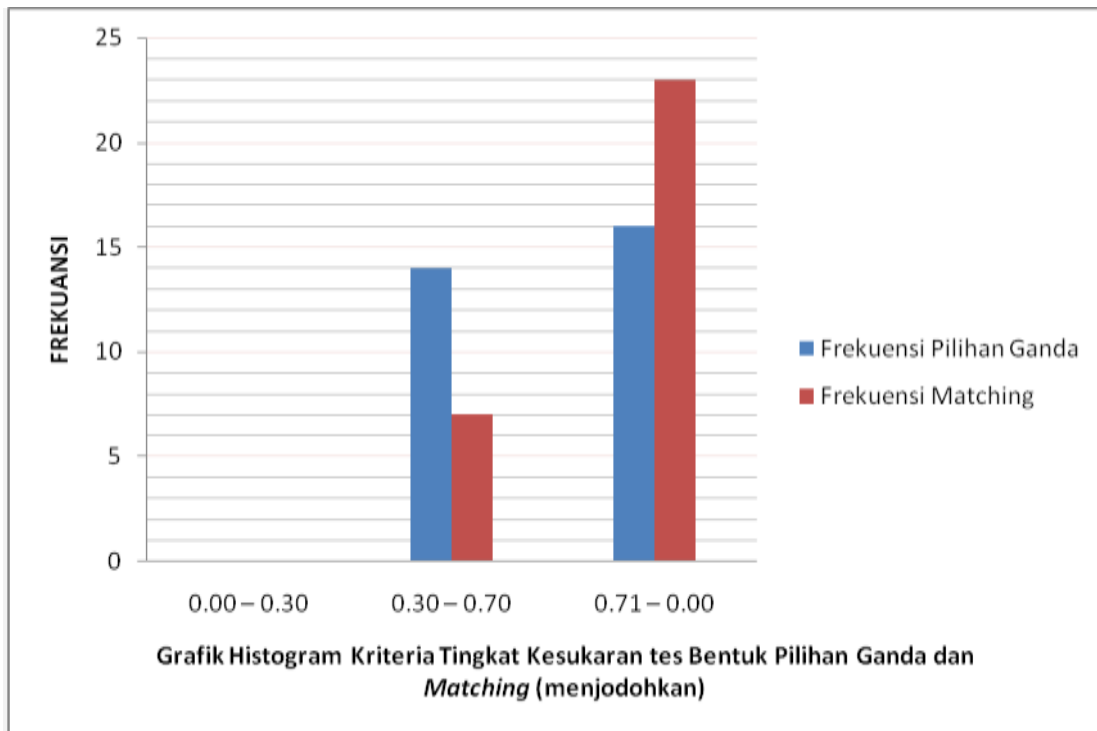
**Tabel 4.5: Tabel Data Distribusi Frekuensi Tingkat Kesukaran Soal pada Tes Bentuk Pilihan Ganda**

Nilai	Frekuensi	Kriteria
0.00 – 0.30	0	Sukar
0.30 – 0.70	7	Sedang
0.71 – 1.00	23	Mudah

**Tabel 4.6: Tabel Data Distribusi Frekuensi Tingkat Kesukaran Soal pada Tes Bentuk Menjodohkan (*matching test*)**

Nilai	Frekuensi	Kriteria
0.00 – 0.30	0	Sukar
0.30 – 0.70	14	Sedang
0.71 – 1.00	16	mudah

Berdasarkan tabel distribusi frekuensi pada masing-masing bentuk tes di atas, maka perbedaan tingkat kesukaran antara tes bentuk pilihan ganda dengan tes menjodohkan (*matching test*) pada Histogram sebagai berikut:



**Gambar 4.7: Diagram Tingkat Kesukaran Soal pada Tes Bentuk Pilihan Ganda dengan Menjodohkan (*Matching Test*)**

Gambar histogram menunjukkan bahwa frekuensi untuk kategori soal sukar pada tes bentuk pilihan ganda dan menjodohkan (*matching test*) adalah sama yaitu 0, kategori soal sedang pada tes bentuk pilihan ganda 7 dan tes menjodohkan (*matching test*) 14, artinya frekuensi untuk kategori soal sedang pada pilihan ganda lebih rendah dibanding dengan tes menjodohkan (*matching test*), sedangkan kategori soal mudah pada tes pilihan ganda adalah 23 dan *matching* 16, artinya frekuensi untuk kategori soal mudah pada bentuk tes pilihan ganda lebih tinggi dibanding dengan tes menjodohkan (*matching test*).

Adapun untuk mengetahui perbedaan rerata tingkat kesukaran butir antara kedua bentuk tes tersebut akan dianalisis menggunakan Analisis uji-t untuk mengetahui sejauh mana perbedaan rerata tingkat kesukaran kedua bentuk tes tersebut.

Rerata tingkat kesukaran antara tes bentuk pilihan ganda dengan bentuk tes menjodohkan (*matching tes*). Sebagaimana dijelaskan pada landasan teoritik bahwa analisis untuk membandingkan perbedaan dua rerata adalah dengan menggunakan analisis uji-t

Untuk menentukan uji-t yang akan dipilih dalam menentukan pengujian hipotesis, maka perlu diuji dulu varians kedua sampel homogen atau tidak. Dan berdasarkan hasil analisis dengan menggunakan uji F, maka menunjukkan bahwa  $f_{hit} = 2,085$  dan  $f_{tab} = 2,10$  pada taraf signifikan  $\alpha = 0,05$ , maka  $f_{hit} < f_{tab}$ , artinya varians tingkat kesukaran kedua bentuk tes yaitu pilihan ganda dengan menjodohkan (*matching test*) adalah homogen.

Berdasarkan hasil analisis data dengan menggunakan uji-t menunjukkan Untuk  $\alpha = 0,05$  dengan  $dk = 30 - 1$  adalah 2,224 sedangkan nilai  $t_{hitung} = 0,224$ , dengan kriteria pengujian dua pihak dimana  $-t_{\alpha/2} < t_{hitung} < t_{\alpha/2}$  maka  $H_0$  diterima dan  $H_1$  ditolak sehingga,  $-2,045 < t_{hitung} < 2,224$  artinya  $H_0$  diterima atau dapat disimpulkan bahwa tidak terdapat perbedaan tingkat kesukaran butir antara tes bentuk pilihan ganda dengan tes menjodohkan (*matching test*) pada mata pelajaran Bahasa Arab pada siswa kelas XI di Madrasah Aliyah Negeri Waimital Kabupaten Seram Bagian Barat.

#### **D. Kesimpulan**

Tingkat kesukaran tes bentuk pilihan ganda tidak lebih tinggi dari pada tes menjodohkan (*matching tes*). Hal ini membenarkan asumsi tingkat kesukaran antara tes pilihan ganda dengan matching memiliki tingkat kesukaran yang sama dimana soal pilihan ganda dengan sedikit alternatif jawaban tetapi tingkat pengecohnya lebih tinggi dan tes matching dengan banyak alternative jawaban tetapi tingkat pengecohnya lebih rendah.

Berdasarkan hasil pengujian hipotesis uji-t didapatkan bahwa tingkat kesukaran pada tes bentuk ganda tidak lebih tinggi dari pada bentuk menjodohkan (*matching tes*). ini membuktikan bahwa beragamnya bentuk soal obyektif, maka tidak bisa memberikan *jagement* bahwa soal pilihan ganda memiliki tingkat kesukaran yang lebih tinggi. Meskipun pada bentuk tes pilihan ganda terdapat dua parameter yang menyebabkan tingginya tingkat kesukaran yaitu adanya pengecoh (*distractor*) dan adanya peluang responden untuk menebak jawaban.



## **Daftar Pustaka**

- Aiken, dalam *Surapranata, Analisis validitas, Reliabilitas, dan Interpretasi Hasil Tes*, Bandung: Remaja Rosdakarya, 2004.
- Anastasi, Anne, *Psychological Testing*, New York: Mac Millan Publishing Company, 1988.
- Arikuto, Suharsimi, *Dasar-dasar Evaluasi Pendidikan*, Jakarta :Bumi Aksara, 1999.
- Bloom, Benjamin S.,Engelhart, M.D., Furst, E.J., Hill, W.H., dan Krathwohl, D.R, *Taxonomy of Educatin Objectivees, Hanbook I: Cognitive Domaian*, New York; David Mc. Kay Company, Inc, 1971.
- Djaali dan Mulyono, *Pengukuran Dalam Bidang Pendidikan*, Jakarta, PT. Gramedia, 2008.
- Gronlund, Norman E., *Measurement and Evaluation in Teaching*. New York: Machmillan Publishing Company, 1985.
- Hamalik, Oemar, *Evaluasi Kurikulum*, Bandung: Remaja Rosdakarya, 1993.
- Naga, Dali S., *Teori Pengukuran*, Jakarta: Program Pasca Sarjana Universitas Negeri Jakarta, 2008.
- Purwanto, *Evaluasi Hasil Belajar*, Yogyakarta: Pustaka Pelajar, 2009.
- Purwanto, Ngalm, *Prinsip-prinsip dan Teknik Evaluasi Pengajaran*, Bandung: PT. Remaja Rosdakarya. 2008.
- Sappaile, Pallawagau, *Pengaruh Tipe Soal Dan Waktu Testing Terhadap Daya Diskriminator Butir Soal Biologi SMU*, Tesis Jakarta: PPS UNJ, 2005.
- Silverius, Suke *Evaluasi Hasil Belajar dan Umpan Balik*, Jakarta: Grasindo, 1991.
- Sukardi, *Evaluasi Pendidikan Prinsip dan Operasionalnya*, Jakarta: Bumi Aksara, 2009.
- Sudjana, Nana, *Penilaian Hasil Proses Belajar Mengajar*, Bandung: PT. Remaja Rosdakarya, 2009.
- , *Dasar-dasar Proses Belajar Mengajar*, Bandung: Sinar Baru Algesindo, 2005.
- Sudijono, Anas, *Pengantar Evaluasi Pendidikan*, Jakarta: Raja Grafindo Persada, 2005.
- Uno, Hamzah B. , Herminanto Sofyan, dan I Made Candiasa, *Pengembangan Instrumen Untuk Penelitian*, Jakarta: Delima Press, 2001.