

## BAB II

### LANDASAN TEORI

#### A. Deskripsi Teori

1. Instrumen Tes
  - a. Pengertian tes

Secara etimologis, istilah “tes” berasal dari bahasa latin “*testum*” yang berarti: sebuah piring atau jambangan dari tanah liat. Dalam pengertian yang luas, tes adalah alat atau instrumen yang dipakai untuk mengukur sesuatu.<sup>1</sup>

Ada beberapa istilah yang berkaitan dengan uraian diatas, yaitu istilah *tes*, *testing*, *testee*, *tester*, yang masing-masing mempunyai pengertian yang berbeda. *Tes* merupakan alat atau prosedur yang digunakan untuk mengetahui atau mengukur sesuatu dalam suasana, dengan cara dan aturan-aturan yang sudah ditentukan. *Testing* merupakan saat pada waktu tes itu dilaksanakan atau dapat juga dikatakan adalah saat pengambilan tes. *Testee* adalah responden yang sedang mengerjakan tes. *Tester* adalah orang yang disertai untuk melaksanakan pengambilan tes terhadap para responden.<sup>2</sup>

---

<sup>1</sup> Shodiq Abdullah, *Evaluasi Pembelajaran: Konsep Dasar, Teori dan Aplikasi*, (Semarang: Pustaka Rizki Putra, 2012), hlm. 1

<sup>2</sup> Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, hlm. 53

Menurut S. Eko Putro Widoyoko, tes adalah diartikan sebagai sejumlah pernyataan yang harus diberikan tanggapan dengan tujuan untuk mengukur tingkat kemampuan seseorang atau mengungkap aspek tertentu dari orang yang dikenai tes.<sup>3</sup>

Menurut Suharsimi Arikunto, tes adalah serentetan pertanyaan atau latihan serta alat lain yang digunakan untuk mengukur keterampilan, pengetahuan, intelegensi, kemampuan atau bakat yang dimiliki oleh individu atau kelompok.<sup>4</sup>

Menurut M. Chabib Thoha, tes adalah alat pengukuran berupa pertanyaan, perintah, dan petunjuk yang ditujukan kepada *testee* untuk mendapatkan respon sesuai dengan petunjuk itu.<sup>5</sup>

Menurut William Wiersma Stephen G.Jurs, dalam bukunya *Educational Measurement and Testing* mendefinisikan “*The test is the stimulus to which the*

---

<sup>3</sup> S. Eko putro Widoyoko, *Evaluasi Program Pembelajaran: Panduan Praktis bagi Pendidik dan Calon Pendidik*, (Yogyakarta: Pustaka Pelajar, 2009), hlm. 45-46

<sup>4</sup> Suharsimi Arikunto, *Prosedur Penelitian Suatu Pendekatan Praktik*, (Jakarta: Rineka Cipta, 2010), hlm. 193

<sup>5</sup> M. Chabib Thoha, *Teknik Evaluasi Pendidikan*, hlm. 43

*response is made*”<sup>6</sup> Artinya tes adalah suatu rangsangan yang membuat orang untuk menanggapi (merespon).

Berdasarkan beberapa pengertian tes di atas dapat disimpulkan bahwa tes adalah suatu alat pengumpul informasi melalui serentetan pertanyaan, perintah atau latihan yang digunakan untuk mengukur keterampilan, pengetahuan, intelegensi, kemampuan atau bakat yang ditunjukkan kepada *testee*.

b. Macam-macam tes

1) Berdasarkan objek pengukurannya, tes dibagi menjadi dua, yaitu:

a) Tes kepribadian

Tes kepribadian adalah tes yang dilaksanakan dengan tujuan mengungkap ciri-ciri khas dari seseorang yang banyak sedikitnya bersifat lahiriah, seperti gaya bicara, cara berpakaian, nada suara, hobi atau kesenangan dan lain-lain.<sup>7</sup> Yang termasuk dalam jenis tes ini adalah tes sikap, tes minat, tes bakat dan tes intelegensi.<sup>8</sup>

---

<sup>6</sup> William Wiersma Stephen G. Jurs, *Educational Measurement and Testing*, (United States: A Division of Simon & Schuster, 1990), hlm. 9

<sup>7</sup> Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 73

<sup>8</sup> M. Chabib Thoha, *Teknik Evaluasi Pendidikan*, hlm. 44

b) Tes hasil belajar

Tes hasil belajar ialah tes yang dipergunakan untuk menilai hasil-hasil pelajaran yang telah diberikan oleh guru kepada murid-muridnya, atau oleh dosen kepada mahasiswanya dalam jangka waktu tertentu.<sup>9</sup>

2) Berdasarkan fungsinya, tes dapat dibedakan dalam empat jenis, yaitu:

a) Tes penempatan

Tes penempatan adalah tes untuk mengukur kemampuan dasar yang dimiliki oleh anak didik, kemampuan tersebut dapat dipakai meramalkan kemampuan peserta didik pada masa mendatang, sehingga kepadanya dapat dibimbing, diarahkan atau ditempatkan pada jurusan yang sesuai dengan kemampuan dasarnya.<sup>10</sup> Pada umumnya tes penempatan dibuat sebagai *pre-test*.<sup>11</sup>

b) Tes formatif

Tes formatif adalah kegiatan penilaian yang bertujuan untuk mencari umpan balik

---

<sup>9</sup> M. Ngalim Purwanto, *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*, hlm. 33

<sup>10</sup> M. Chabib Thoaha, *Teknik Evaluasi Pendidikan*, hlm. 46

<sup>11</sup> Zainal Arifin, *Evaluasi Pembelajaran: Prinsip Teknik Prosedur*, hlm. 36

(*feedback*), yang selanjutnya hasil penilaian tersebut dapat digunakan untuk memperbaiki proses belajar mengajar yang sedang atau yang sudah dilaksanakan. Jadi, sebenarnya tes formatif itu tidak hanya dilakukan pada tiap akhir pelajaran, tetapi bisa juga ketika pelajaran sedang berlangsung.<sup>12</sup>

c) Tes diagnostik

Tes diagnostik adalah suatu tes yang digunakan untuk mengetahui kekurangan-kekurangan siswa sehingga berdasarkan kekurangan-kekurangan tersebut dapat dilakukan pemberian perlakuan yang sesuai.<sup>13</sup>

d) Tes sumatif

Tes sumatif adalah penilaian yang dilakukan untuk memperoleh data atau informasi sampai di mana penguasaan atau pencapaian belajar siswa terhadap bahan pelajaran yang telah dipelajarinya selama jangka waktu tertentu.<sup>14</sup>

---

<sup>12</sup> M. Ngalim Purwanto, *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*, hlm. 26

<sup>13</sup> Ign Masidjo, *Penilaian Pencapaian Hasil Belajar Siswa di Sekolah*, (Yogyakarta: Kanisius, 1995), hlm. 54

<sup>14</sup> M. Ngalim Purwanto, *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*, hlm. 26

Tes sumatif ini di sekolah sering dikenal dengan istilah “Ulangan Umum” atau ”EBTA” (Evaluasi Belajar Tahap Akhir), di mana hasilnya digunakan untuk mengisi nilai rapor atau mengisi ijazah (STTB).<sup>15</sup>

3) Berdasarkan tingkatannya,

a) Tes standar

Tes standar adalah tes yang telah mengalami proses standardisasi, yakni proses validasi dan keandalan (*reliability*) sehingga tes tersebut benar-benar valid dan andal untuk suatu tujuan dan bagi suatu kelompok tertentu.<sup>16</sup>

b) Tes nonstandar

Tes nonstandar adalah tes yang disusun oleh seorang pendidik yang belum memiliki keahlian profesional dalam penyusunan tes, atau mereka yang memiliki keahlian tetapi tidak sempat menyusun tes secara baik, mengujicobakan, melakukan analisis sehingga validitas dan reliabilitas belum dapat dipertanggungjawabkan.<sup>17</sup>

---

<sup>15</sup> Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 72

<sup>16</sup> M. Ngalim Purwanto, *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*, hlm. 33

<sup>17</sup> M. Chabib Thoha, *Teknik Evaluasi Pendidikan*, hlm. 52

4) Berdasarkan bentuknya

a) Tes tertulis

Tes tertulis adalah jenis tes di mana tester dalam mengajukan butir-butir pertanyaan atau soalnya dilakukan secara tertulis dan *testee* memberikan jawabannya juga secara tertulis.<sup>18</sup>

Tes tertulis secara umum dapat dibedakan menjadi dua bagian, yaitu:

(1) Tes obyektif

Tes obyektif yang juga dikenal dengan istilah tes jawaban pendek, tes “ya-tidak” dan tes model baru, adalah salah satu jenis tes hasil belajar yang terdiri dari butir-butir soal (items) yang dapat dijawab oleh *testee* dengan jalan memilih salah satu (atau lebih) di antara beberapa kemungkinan jawaban yang telah dipasangkan pada masing-masing items atau dengan jalan menuliskan (mengisikan) jawaban berupa kata-kata atau simbol-simbol tertentu pada tempat atau ruang yang telah disediakan

---

<sup>18</sup> Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 75

untuk masing-masing butir item yang bersangkutan.<sup>19</sup>

Bentuk tes obyektif ada beberapa macam, yaitu:

(a) *Free response item* (tes jawab bebas), terdiri atas:

(i) *Completion test* (tes melengkapi)

Tes completion merupakan salah-satu bentuk tes jawaban bebas, di mana butir-butir soalnya berupa satu kalimat di mana bagian-bagian tertentu yang dianggap penting dikosongkan.

(ii) *Fill-in* (mengisi titik-titik dalam kalimat yang dikosongkan)

Tes obyektif bentuk fill in (bentuk isian) ini biasanya berbentuk cerita atau karangan. Kata-kata penting dalam cerita atau karangan itu beberapa diantaranya dikosongkan (tidak dinyatakan), sedangkan tugas *testee* adalah

---

<sup>19</sup> Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 106-107

mengisi bagian-bagian yang telah dikosongkan itu.<sup>20</sup>

(b) *Selection type test* (tes yang menjawab dengan mengadakan pilihan), yang terdiri atas:

(i) Tes benar salah (*True-False*)

Tes tipe benar salah adalah tes yang butir soalnya terdiri dari pernyataan yang disertai dengan alternatif jawaban yaitu jawaban atau pernyataan yang benar dan yang salah. Peserta tes diminta untuk menandai masing-masing jawaban atau pernyataan itu dengan melingkari ataupun memberi tanda silang pada huruf “B” jika jawaban atau pernyataan itu dianggap benar menurut pendapatnya dan melingkari ataupun memberi tanda silang pada huruf “S” jika jawaban atau pernyataan itu menurut pendapatnya dianggap salah.<sup>21</sup>

---

<sup>20</sup> Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 114

<sup>21</sup> S. Eko Putro Widoyoko, *Evaluasi Program Pembelajaran: Panduan Praktis Bagi Pendidik dan Calon Pendidik*, hlm. 51

(ii) Tes menjodohkan (*Matching*)

Ada beberapa istilah yang digunakan untuk menunjuk tes menjodohkan, seperti memasangkan atau mencocokkan. Tes ini terdiri dari satu seri pertanyaan dan satu seri jawaban. Tugas *testee* adalah mencari dan menempatkan jawaban-jawaban yang telah tersedia, sehingga sesuai atau cocok atau merupakan pasangan atau merupakan jodoh dari pertanyaannya.<sup>22</sup>

(iii) Tes pilihan ganda (*Multiple Choice*)

Tes pilihan ganda adalah tes yang terdiri atas suatu keterangan atau pemberitahuan tentang suatu pengertian yang belum lengkap. Dan untuk melengkapinya harus memilih satu

---

<sup>22</sup> Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 111

dari beberapa kemungkinan jawaban yang telah disediakan.<sup>23</sup>

(2) Tes subyektif (uraian)

Tes subyektif, pada umumnya berbentuk uraian (esai). Tes bentuk uraian adalah pertanyaan yang menuntut siswa menjawabnya, mendiskusikan, membandingkan, memberikan alasan, dan bentuk lain yang sejenis sesuai dengan tuntutan pertanyaan dengan menggunakan kata-kata dan bahasa sendiri.<sup>24</sup>

b) Tes lisan

Tes lisan adalah tes di mana tester di dalam mengajukan pertanyaan-pertanyaan atau soalnya dilakukan secara lisan, dan *testee* memberikan jawabannya secara lisan pula.<sup>25</sup> Dilihat dari segi persiapan dan cara bertanya tes lisan dapat dibedakan menjadi dua, yaitu<sup>26</sup>:

---

<sup>23</sup> Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, hlm. 168

<sup>24</sup> Nana Sudjana, *Penilaian Hasil Proses Belajar Mengajar*, (Bandung: PT Remaja Rosdakarya, 2010), hlm. 35

<sup>25</sup> Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 75

<sup>26</sup> M. Chabib Thoha, *Teknik Evaluasi Pendidikan*, hlm. 61

(1) Tes lisan bebas

Tes lisan bebas artinya pendidik dalam memberikan soal kepada peserta didik tanpa menggunakan pedoman yang dipersiapkan secara tertulis.

(2) Tes lisan berpedoman

Tes lisan berpedoman artinya pendidik menggunakan pedoman tertulis tentang apa yang akan ditanyakan kepada peserta didik.

c) Tes tindakan

Tes tindakan adalah tes di mana respon atau jawaban yang dituntut dari peserta didik berupa tindakan, tingkah laku kongkrit. Alat yang dapat digunakan untuk melakukan tes ini adalah observasi atau pengamatan terhadap tingkah laku tersebut.<sup>27</sup>

c. Kriteria tes yang baik

Menurut Suharsimi Arikunto, suatu tes dapat dikatakan baik apabila memenuhi lima persyaratan, yaitu: validitas, reliabilitas, obyektivitas, praktikabilitas dan ekonomis.<sup>28</sup>

1) Validitas

Validitas merupakan syarat yang terpenting dalam suatu alat evaluasi. Suatu tes disebut valid

---

<sup>27</sup> M. Chabib Thoah, *Teknik Evaluasi Pendidikan*, hlm. 63

<sup>28</sup> Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, hlm.57-58

apabila tes tersebut dapat mengukur apa yang hendak dan seharusnya diukur.<sup>29</sup> Jadi, tes hasil belajar dapat dinyatakan valid apabila tes hasil belajar tersebut (sebagai alat pengukur keberhasilan dapat belajar peserta didik) dengan secara tepat, benar, shahih atau absah telah dapat mengukur atau mengungkap hasil-hasil belajar yang telah dicapai oleh peserta didik, setelah mereka menempuh proses belajar mengajar dalam jangka waktu tertentu.<sup>30</sup>

## 2) Reliabilitas.

Reliabilitas adalah *ketetapan* atau *ketelitian* suatu alat evaluasi. Suatu tes atau alat evaluasi dikatakan andal jika ia dapat dipercaya, konsisten, atau stabil dan produktif.<sup>31</sup> Tes dikatakan dapat dipercaya (*reliable*) jika memberikan hasil yang tetap atau *ajeg* (*consistent*) apabila diteskan berkali-kali. Jika kepada siswa diberikan tes yang sama pada waktu yang berlainan, maka setiap siswa akan tetap berada dalam urutan (rangking) yang sama atau ajek dalam kelompoknya.<sup>32</sup>

---

<sup>29</sup> Mudjijo, *Tes Hasil Belajar*, (Jakarta: Bumi Aksara, 1995), hlm. 40

<sup>30</sup> Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 93-94

<sup>31</sup> M. Ngalim Purwanto, *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*, hlm. 139

<sup>32</sup> S. Eko Putro Widoyoko, *Evaluasi Program Pembelajaran*, hlm. 99

### 3) Obyektivitas

Objektif berarti tidak adanya unsur pribadi yang mempengaruhinya. Lawan dari objektif adalah subjektif, artinya terdapat unsur pribadi yang masuk memengaruhinya. Sebuah tes dikatakan memiliki obyektivitas apabila dalam melaksanakan tes tidak ada faktor subjektif yang memengaruhinya, terutama dalam sistem skoringnya.

Ada 2 faktor yang memengaruhi subjektivitas dari suatu tes, yaitu bentuk tes dan penilai. Bentuk tes uraian akan memberi banyak kemungkinan kepada penilai untuk memberikan penilaian menurut caranya sendiri. Dengan demikian maka hasil dari seorang siswa yang mengerjakan soal dari sebuah tes, akan memperoleh skor yang berbeda apabila dinilai oleh dua orang. Itulah sebabnya pada waktu sekarang ini ada kecenderungan penggunaan tes objektif di berbagai bidang. Untuk menghindari masuknya unsur subjektivitas dari penilai, maka sistem skoringnya dapat dilakukan dengan sebaik-baiknya, antara lain dengan membuat pedoman skoring terlebih dahulu.<sup>33</sup>

---

<sup>33</sup> S. Eko Putro Widoyoko, *Evaluasi Program Pembelajaran: Panduan Praktis Bagi Pendidik dan Calon Pendidik*, hlm. 100

#### 4) Praktikabilitas.

Sebuah tes dikatakan memiliki praktikabilitas yang tinggi apabila tes tersebut bersifat praktis, mudah pengadministrasiannya. Tes yang praktis adalah tes yang:

- a) Mudah dilaksanakan, artinya tidak menuntut peralatan yang banyak dan memberi kebebasan kepada siswa untuk mengerjakan terlebih dahulu bagian yang dianggap mudah oleh siswa.
- b) Mudah pemeriksaannya, artinya bahwa tes itu dilengkapi dengan kunci jawaban maupun pedoman skoringnya. Untuk soal bentuk obyektif, pemeriksaan akan lebih mudah dilakukan jika dikerjakan oleh siswa dalam lembar jawaban.
- c) Dilengkapi dengan petunjuk-petunjuk sehingga dapat diberikan oleh orang lain.

#### 5) Ekonomis

Ekonomis di sini adalah bahwa pelaksanaan tes tersebut tidak membutuhkan biaya yang mahal, tenaga yang banyak dan waktu yang lama.<sup>34</sup>

---

<sup>34</sup> S. Eko Putro Widoyoko, *Evaluasi Program Pembelajaran: Panduan Praktis Bagi Pendidik dan Calon Pendidik*, hlm. 101-102

## 2. Validitas Tes

### a. Pengertian validitas

Validitas berasal dari kata *validity*, yang dapat diartikan sebagai *ketepatan* atau *kesahihan*, yaitu sejauh mana sebuah instrumen atau alat ukur mampu atau berhasil mengukur apa yang seharusnya diukur.<sup>35</sup>

Menurut M. Chabib Thoha dalam bukunya *teknik evaluasi pendidikan*, suatu alat ukur disebut memiliki validitas bilamana alat ukur tersebut isinya layak mengukur obyek yang seharusnya diukur dan sesuai dengan kriteria tertentu. Artinya adanya kesesuaian antara alat ukur dengan fungsi pengukuran dan sasaran pengukuran.<sup>36</sup>

Menurut Lewis R. Aiken dalam bukunya *Psychological Testing and Assessment*, mendefinisikan validitas tes: *The validity of a test has been defined as the extent to which the test measures what is was designed to measure.*<sup>37</sup> Validitas suatu tes diartikan sebagai sejauh mana tes tersebut mengukur apa yang hendak diukur.

---

<sup>35</sup> Shodiq Abdullah, *Evaluasi Pembelajaran: Konsep Dasar, Teori dan Aplikasi*, hlm. 76

<sup>36</sup> M. Chabib Thoha, *Teknik Evaluasi pendidikan*, hlm. 109-110

<sup>37</sup> Lewis R. Aiken, *Psychological Testing and Assessment*, (America: ISBN, 1991), hlm. 105

Menurut Muhammad Abdul Khalik Muhammad dalam kitabnya *Ikhtibaarootun al-Lughah*, mendefinisikan validitas adalah sebagai berikut:

ان صدق الاختبار يعني الى اي مدى يقيس الاختبار الشيء الذي وضع من اجله<sup>38</sup>

“Validitas tes adalah sejauh mana tes tersebut dapat mengukur apa-apa yang hendak diukur.”

Menurut Mudjjo, validitas adalah tingkat ketetapan tes tersebut dalam mengukur materi dan perilaku yang harus diukur.<sup>39</sup>

Menurut Sumarna Surapranata, validitas adalah suatu konsep yang berkaitan dengan sejauh mana tes telah mengukur apa yang seharusnya diukur.<sup>40</sup>

Berdasarkan berbagai pengertian di atas, dapat disimpulkan bahwa tes hasil belajar dapat dinyatakan valid apabila tes hasil belajar tersebut (sebagai alat pengukur keberhasilan belajar peserta didik) dengan secara tepat, benar, shahih atau absah telah dapat mengukur atau mengungkap hasil-hasil belajar yang telah

---

<sup>38</sup> Muhammad Abdul Khalik Muhammad, *Iktibatul al-Lughah*, (Riyad: Jami'ah Malik Su'ud, 1989), hlm.48

<sup>39</sup> Mudjjo, *Tes Hasil Belajar*, hlm. 40

<sup>40</sup> Sumarna Surapranata, *Analisis, Validitas, Reliabilitas, dan Interpretasi Hasil Tes*, (Bandung: PT Remaja Rosdakarya, 2004), hlm. 50

dicapai oleh peserta didik, setelah mereka menempuh proses belajar mengajar dalam jangka waktu tertentu.

b. Macam-macam validitas

Secara garis besar ada 2 (dua) macam validitas, yaitu validitas soal dan validitas butir soal.

1) Validitas tes

Secara garis besar ada dua macam validitas, yaitu validitas logis dan validitas empiris.

a) Validitas logis

Istilah “validitas logis” mengandung kata “logis” berasal dari kata “logika” yang berarti penalaran. Dengan makna demikian maka validitas logis untuk sebuah instrumen evaluasi menunjuk pada kondisi bagi sebuah instrumen yang memenuhi persyaratan valid berdasarkan hasil penalaran.<sup>41</sup>

Ada dua macam validitas logis yang dapat dicapai oleh sebuah instrumen, yaitu validitas isi dan validitas konstruksi.

(1) Validitas isi

Validitas isi adalah validitas yang ditilik dari segi isi tes itu sendiri sebagai alat pengukur hasil belajar, yaitu: sejauh mana tes hasil belajar sebagai alat pengukur hasil

---

<sup>41</sup> Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, hlm. 65

belajar peserta didik, isinya telah dapat mewakili secara representatif terhadap keseluruhan materi atau bahan pelajaran yang seharusnya diteskan (diujikan).<sup>42</sup>

Pengujian validitas isi yang dilakukan dengan menelaah butir (*item review*) dilakukan dengan mencermati kesesuaian isi butir yang ditulis dengan perencanaan yang dituangkan dalam kisi-kisi. Butir-butir tes dinyatakan valid (*logically valid*) apabila setelah mencermati isi butir-butir yang ditulis telah menunjukkan kesesuaian dengan kisi-kisi.<sup>43</sup>

## (2) Validitas konstruksi

Validitas konstruksi adalah suatu tes di mana butir soal tersebut membangun setiap aspek berfikir seperti yang disebutkan dalam tujuan instruksional khusus.<sup>44</sup> Menurut Benjamin S. Bloom bahwa taksonomi (pengelompokan) tujuan pendidikan harus senantiasa mengacu kepada tiga jenis ranah (*domain*) yang melekat pada diri peserta

---

<sup>42</sup> Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 164

<sup>43</sup> Purwanto, *Evaluasi Hasil Belajar*, hlm. 120-121

<sup>44</sup> M. Chabib Thoha, *Teknik Evaluasi Pendidikan*, hlm. 110

didik, yaitu: Ranah proses berpikir (*cognitive domain*), Ranah nilai atau sikap (*affective domain*), Ranah keterampilan (*psychomotor domain*). Setiap domain disusun menjadi beberapa jenjang kemampuan, mulai dari hal yang sederhana sampai dengan hal yang kompleks, mulai dari hal yang mudah sampai dengan hal yang sukar, dan mulai dari hal yang konkrit sampai dengan hal yang abstrak. Adapun rincian domain tersebut adalah sebagai berikut:

(a) Ranah kognitif

Benjamin S. Bloom membagi tingkat kemampuan atau tipe hasil belajar yang termasuk aspek kognitif menjadi enam, yaitu:

i. Pengetahuan (*knowledge*)

Tingkat kemampuan yang hanya meminta responden atau *testee* untuk mengenal atau mengetahui adanya konsep, fakta, atau istilah-istilah tanpa harus dimengerti, atau dapat menilai, atau dapat menggunakan. Kata kerja operasional-nya, antara lain: menyebutkan, menunjuk-

kan, mengenal, mengingat kembali, mendefinisikan.

ii. Pemahaman (*komprehensif*)

Tingkat kemampuan yang diharapkan testee mampu memahami arti atau konsep, situasi, serta fakta yang diketahuinya. Kata kerja operasional yang dipakai pada ranah ini adalah: membedakan, mengubah, mempersiapkan, menyajikan, mengatur, menginterpretasikan, menjelaskan, mendemonstrasikan, memberi contoh, memperkirakan, menentukan, mengambil kesimpulan.

iii. Penerapan (*aplikasi*)

Dalam tingkat aplikasi, *testee* atau responden dituntut kemampuannya untuk menerapkan atau menggunakan apa yang telah diketahuinya dalam suatu situasi yang baru baginya. Kata kerja operasional untuk ranah ini adalah: menggunakan, menerapkan, menggeneralisasikan, menghubungkan, memilih, mengembangkan, mengorganisasi, menyusun,

mengklasifikasikan, mengubah struktur.

iv. Analisis

Tingkat kemampuan *testee* untuk menganalisis atau menguraikan suatu integritas atau suatu situasi tertentu ke dalam komponen-komponen atau unsur-unsur pembentuknya. Pada tingkat analisis, *testee* diharapkan dapat memahami dan sekaligus dapat memilah-milahnya menjadi bagian-bagian. Kata kerja untuk ranah ini adalah membedakan, menemukan, mengklasifikasikan, mengategorikan, membandingkan, mengadakan pemisahan.

v. Sintesis

Yang dimaksud dengan sintesis ialah penyatuan unsur-unsur atau bagian-bagian ke dalam suatu bentuk yang menyeluruh. Dengan kemampuan sintesis seseorang dituntut untuk dapat menemukan hubungan kausal atau urutan tertentu, atau menemukan abstraksinya yang berupa

integritas. Kata kerja operasional untuk ranah ini adalah menghubungkan, menghasilkan, mengkhususkan, mengembangkan, menggabungkan, mengorganisasi, menyintesis, mengklasifikasikan, menyimpulkan.

vi. Evaluasi

Dengan kemampuan evaluasi, *testee* diminta untuk membuat suatu penilaian tentang suatu pernyataan, konsep, situasi. Kata kerja operasional untuk ranah ini adalah menafsirkan, menilai, menentukan, mempertimbangkan, membandingkan, melakukan, memutuskan, mengargumentasikan, menaksir.<sup>45</sup>

(b) Ranah afektif

Ranah afektif adalah ranah yang berkaitan dengan sikap dan nilai. Ciri-ciri hasil belajar afektif akan tampak pada peserta didik dalam berbagai tingkah laku, seperti perhatiannya terhadap mata

---

<sup>45</sup> M. Ngalim Purwanto, *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*, hlm. 43-47

pelajaran pendidikan agama Islam, kedisiplinannya dalam mengikuti pelajaran agama di sekolah, motivasinya yang tinggi untuk tahu lebih banyak mengenai pelajaran agama Islam yang diterimanya, penghargaan atau rasa hormatnya terhadap guru pendidikan Agama Islam. Ranah afektif di taksonomi menjadi lebih rinci lagi ke dalam lima jenjang: *receiving, responding, valuing, organization, characterization by a value or value complex*.

- i. *receiving* atau *attending* (menerima atau memperhatikan), adalah kepekaan seseorang dalam menerima rangsangan (stimulus) dari luar yang datang dari luar.
- ii. *responding* (menanggapi) mengandung arti “adanya partisipasi aktif”. Jadi kemampuan menanggapi adalah kemampuan yang dimiliki oleh seseorang untuk mengikutsertakan dirinya secara aktif dalam fenomena tertentu dan membuat reaksi terhadapnya dengan salah satu cara.

- iii. *valuing* (menilai = menghargai).  
Menilai atau menghargai artinya memberikan nilai atau memberikan penghargaan terhadap suatu kegiatan atau obyek, sehingga apabila kegiatan itu tidak dikerjakan, dirasakan akan membawa kerugian atau penyesalan.
- iv. *organization* (mengatur atau mengorganisasikan) artinya menemukan perbedaan nilai sehingga terbentuk nilai baru yang lebih universal, yang membawa kepada perbaikan umum.
- v. *characterization by a value or value complex*, (karakterisasi dengan suatu nilai atau kompleks nilai), yakni keterpaduan semua sistem nilai yang telah dimiliki seseorang, yang mempengaruhi pola kepribadian atau tingkah lakunya.<sup>46</sup>

(c) Ranah psikomotorik

Ranah psikomotor adalah ranah yang berkaitan dengan keterampilan (*skill*) atau kemampuan bertindak setelah

---

<sup>46</sup> Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 54-56

seseorang menerima pengalaman belajar tertentu.<sup>47</sup> Ranah ini juga diklasifikasi menjadi 7 sub-ranah, yaitu:

i. Persepsi (*Perception*)

Sub ranah ini merujuk pada penggunaan organ-organ indrawi peserta didik untuk mendapatkan gambaran atau kunci yang dapat membimbing gerak atau aktivitas motorik.

ii. Kesiapan (*set*)

Sub ranah ini berkaitan dengan kesiapan untuk mengambil tindakan tertentu, baik dalam bentuk kesiapan mental, fisik, maupun emosional.

iii. Respon terbimbing (*guided response*)

Respon terbimbing merupakan tahap awal dari belajar terhadap keterampilan yang kompleks. Respon yang demikian ini meliputi imitasi (mengulangi perbuatan yang ditunjukkan oleh guru).

---

<sup>47</sup> Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 57

iv. Gerakan terbiasa (*mechanism*)

Sub ranah ini berkaitan dengan tindakan-tindakan di mana respon yang dipelajari telah menjadi kebiasaan, dan gerakan-gerakannya dilakukan dengan percaya diri.

v. Respon kompleks (*complex overt response*)

Pada tahap ini peserta didik telah mampu melakukan gerakan atau aktivitas dengan mahir yang meliputi pola-pola gerak yang kompleks. Profisiensi dalam sub ranah ini diindikasikan dengan kinerja cepat, luwes (tidak kaku), dan akurat dengan mengeluarkan sedikit energi.

vi. Adaptasi (*adaptation*)

Adaptasi berkaitan dengan berbagai ketrampilan yang telah dikembangkan dengan bagus sehingga peserta didik mampu memodifikasi pola-pola gerakan yang dilakukan, untuk disesuaikan dengan situasi atau problem yang dihadapi.

vii. Originasi (*origination*)

Originasi merujuk pada penciptaan pola-pola gerakan atau pola kemahiran baru untuk diterapkan dalam situasi khusus atau problem yang khusus. Hasil belajar dalam sub ranah ini ditunjukkan dengan adanya kreativitas sempurna dan orisinal yang didasarkan atas ketrampilan tingkat tinggi.<sup>48</sup>

b) Validitas empiris

Istilah “validitas empiris” memuat kata “empiris” yang artinya “pengalaman”. Sebuah instrumen dapat dikatakan memiliki validitas empiris apabila sudah diuji dari pengalaman.<sup>49</sup>

Untuk dapat menentukan apakah tes hasil belajar sudah memiliki validitas empirik ataukah belum, dapat dilakukan penelusuran dari dua segi, yaitu dari segi daya ketepatan meramalnya dan daya ketepatan bandingannya.<sup>50</sup>

---

<sup>48</sup> Shodiq Abdullah, *Evaluasi Pembelajaran: Konsep Dasar, Teori, dan Aplikasi*, hlm. 35-38

<sup>49</sup> Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, 66

<sup>50</sup> Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 168

### (1) Validitas ramalan

Validitas ramalan dari suatu tes adalah suatu kondisi yang menunjukkan seberapa jauhkah sebuah tes telah dapat dengan secara tepat menunjukkan kemampuannya untuk meramalkan apa yang bakal terjadi pada masa mendatang.<sup>51</sup>

Untuk mengetahui apakah suatu tes hasil belajar dapat dinyatakan sebagai tes yang telah memiliki validitas ramalan atau belum, dapat ditempuh dengan cara mencari korelasi antara tes hasil belajar yang sedang diuji validitas ramalannya dengan kriterium yang ada. jika diantara kedua variabel tersebut terdapat korelasi positif yang signifikan, maka tes hasil belajar yang sedang diuji validitas ramalannya itu, dapat dinyatakan sebagai tes hasil belajar yang telah memiliki daya ramal yang tepat, artinya apa yang telah diramalkan, betul-betul telah terjadi secara nyata dalam praktek.<sup>52</sup>

---

<sup>51</sup> Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 169

<sup>52</sup> Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 170

## (2) Validitas bandingan

Tes sebagai alat pengukur dapat dikatakan telah memiliki validitas bandingan apabila tes tersebut dalam kurun waktu yang sama dengan secara tepat telah mampu menunjukkan adanya hubungan yang searah, antara tes pertama dengan tes berikutnya. Validitas bandingan juga dikenal dengan istilah validitas sama saat, validitas pengalaman atau validitas ada sekarang. Dikatakan sama saat, sebab validitas tes itu ditentukan atas dasar data hasil tes yang pelaksanaannya dilakukan pada kurun waktu yang sama. Dikatakan validitas pengalaman, sebab validitas tes tersebut ditentukan atas dasar pengalaman yang telah diperoleh. Adapun dikatakan sebagai validitas ada sekarang, sebab setiap kali kita menyebut istilah pengalaman, maka istilah itu akan selalu kita kaitkan dengan hal-hal yang telah ada atau hal-hal yang telah terjadi pada waktu yang lalu, sehingga data mengenai

pengalaman masa lalu itu pada saat sekarang ini sudah ada di tangan.<sup>53</sup>

## 2) Validitas butir tes

Validitas butir (item) dari suatu tes adalah ketepatan mengukur yang dimiliki oleh sebuah item (yang merupakan bagian tak terpisahkan dari tes sebagai suatu totalitas), dalam mengukur apa yang seharusnya diukur lewat butir item tersebut.<sup>54</sup>

Setiap butir item yang ada dalam tes hasil belajar itu merupakan bagian tak terpisahkan dari tes hasil belajar tersebut sebagai suatu totalitas. Eratnya hubungan antara item dengan tes hasil belajar sebagai suatu totalitas itu kiranya dapat dipahami dari kenyataan, bahwa semakin banyak butir-butir item yang dapat dijawab dengan betul oleh *testee*, maka skor-skor total hasil tes tersebut akan semakin tinggi. Sebaliknya, semakin sedikit butir-butir item yang dapat dijawab dengan betul oleh *testee*, maka skor-skor total hasil tes itu akan semakin rendah atau semakin menurun.<sup>55</sup>

---

<sup>53</sup> Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm.176-177

<sup>54</sup> Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 182

<sup>55</sup> Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 182

c. Teknik pengujian validitas item

Sebutir item dapat dikatakan telah memiliki validitas yang tinggi atau dapat dikatakan valid, jika skor-skor pada butir item yang bersangkutan memiliki kesesuaian atau kesejajaran arah dengan skor totalnya, atau dengan bahasa statistiknya: ada korelasi positif yang signifikan antara skor item dengan skor totalnya. Skor total di sini berkedudukan sebagai variabel terikat (*dependent variable*), sedangkan skor item berkedudukan sebagai variabel bebasnya (*independent variable*). Dengan demikian, maka untuk sampai pada kesimpulan bahwa item-item yang ingin diketahui validitasnya, yaitu valid atukah tidak, kita dapat menggunakan teknik korelasi sebagai teknik analisisnya. Sebutir item dapat dinyatakan valid, apabila skor item yang bersangkutan terbukti mempunyai korelasi positif yang signifikan dengan skor totalnya.

Pada tes obyektif maka hanya ada dua kemungkinan jawaban, yaitu betul dan salah. Setiap butir soal yang dijawab dengan betul umumnya diberi skor 1 (satu), sedangkan untuk setiap jawaban salah diberikan skor 0 (nol). Jenis data seperti ini, yaitu betul-salah, ya-tidak atau yang sejenis dengan itu, dalam dunia ilmu statistik dikenal dengan nama data diskrit murni atau data dikotomik. Sedangkan skor total yang dimiliki oleh

masing-masing individu *testee* merupakan hasil penjumlahan dari setiap skor yang dimiliki oleh masing-masing butir item merupakan data kontinyu.

Menurut teori yang ada, apabila variable I berupa data diskrit murni atau data dikotomik, sedangkan variable II berupa data kontinyu, maka teknik korelasi yang tepat untuk digunakan dalam mencari korelasi antara variabel I dengan variabel II adalah teknik korelasi point biserial.<sup>56</sup>

### 3. Reliabilitas Tes

#### a. Pengertian reliabilitas

Kata reliabilitas dalam bahasa Indonesia diambil dari kata *reliability* dalam bahasa Inggris, berasal dari kata asal *reliable* yang artinya dapat dipercaya.<sup>57</sup>

Nana Sudjana dalam bukunya *penilaian Hasil Proses belajar Mengajar*, mengartikan reliabilitas dengan ketetapan atau keajegan alat tersebut dalam menilai apa yang dinilainya. Artinya, kapanpun alat penilaian tersebut digunakan akan memberikan hasil yang relatif sama.<sup>58</sup>

Menurut Anne Anastasi dalam bukunya *Psychological Testing*, mendefinisikan “*Reliability refers*

---

<sup>56</sup> Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 184-185

<sup>57</sup> S. Eko Putro Widoyoko, *Evaluasi Program Pembelajaran: Panduan Praktis Bagi Pendidik dan Calon Pendidik*, hlm. 144

<sup>58</sup> Nana Sudjana, *Penilaian Hasil Proses Belajar Mengajar*. hlm. 16

*to the consistency of scores obtained by the same persons when reexamined with the same test on different occasions, or with different sets of equivalent items, or under other variable examining conditions”.*<sup>59</sup> Artinya, reliabilitas adalah *consistency* atau keajegan atau ketetapan dari nilai yang diperoleh dari tiap individu yang sama manakala diadakan tes ulang dengan tes yang sama pada waktu yang berbeda atau dengan butir soal yang sejenis.

Menurut Muhammad Abdul Malik Muhammad dalam kitabnya *Iktibaarotun al-Lughah*, mendefinisikan reliabilitas tes adalah sebagai berikut:

الثبات: يقصد عدم التذبذب في الاختبار اذا ما قصد به ان يكون  
بمنابة المقياس<sup>60</sup>

“Reliabilitas tes adalah tidak adanya perubahan-perubahan dalam tes yang dilaksanakan dengan menggunakan tes yang serupa”

Berdasarkan beberapa definisi di atas, dapat diambil kesimpulan bahwa suatu tes memiliki keterandalan atau reliabilitas, bilamana tes tersebut dipakai mengukur berulang-ulang hasilnya sama.

---

<sup>59</sup> Anne Anastasi, *Psychological Testing*, (New York: Macmillan Publishing Co, 1982), hlm. 102

<sup>60</sup> Muhammad Abdul Khalik Muhammad, *Iktibatul al-Lughah*, hlm.39

b. Faktor-faktor yang mempengaruhi reliabilitas

Menurut M. Ngalim Purwanto dalam bukunya *prinsip-prinsip dan teknik evaluasi pengajaran*, faktor yang mempengaruhi reliabilitas instrumen evaluasi diantaranya ialah sebagai berikut

1) Luas tidaknya sampling yang diambil

Makin luas suatu sampling, berarti tes makin andal.

2) Perbedaan bakat dan kemampuan murid yang dites

Makin variabel kemampuan peserta tes, berarti makin tinggi keandalan koefisien tes. Tes yang diberikan kepada beberapa tingkat kelas yang berbeda lebih tinggi keandalannya daripada yang hanya diberikan kepada beberapa kelas yang sama karena tingkat kelas yang berbeda akan menghasilkan *achievement* yang lebih luas.

3) Suasana dan kondisi testing

Suasana ketika berlangsung seperti tenang, gaduh, banyak gangguan, pengetes, yang marah-marah dapat mengganggu pengerjaan tes sehingga dengan demikian mempengaruhi pula hasil dan keandalan tes.<sup>61</sup>

---

<sup>61</sup> M Ngalim Purwanto, *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*, hlm.141

c. Teknik pengujian reliabilitas

Dalam rangka menentukan apakah tes hasil belajar bentuk obyektif yang disusun oleh tester telah memiliki keajegan mengukur ataukah belum, dapat dilakukan dengan menggunakan tiga teknik yang dapat digunakan untuk menguji tingkat reliabilitas butir tes, yaitu:

1) Menggunakan pendekatan *single test-single trial* (*single test-single trial method*).

Dalam menentukan reliabilitas tes hasil belajar bentuk obyektif dengan menggunakan pendekatan *single test-single trial*, maka penentuan reliabilitas tes tersebut dilakukan dengan jalan melakukan pengukuran terhadap satu kelompok subyek, di mana pengukuran itu dilakukan dengan hanya menggunakan satu jenis alat pengukur dan bahwa pelaksanaan pengukuran itu hanya dilakukan sebanyak satu kali saja. Dengan kata lain, pendekatan *single test-single trial* adalah merupakan pendekatan “serba single” atau pendekatan “serba satu” yaitu: satu kelompok subyek, satu jenis alat pengukur, dan satu kali pengukuran atau satu kelompok *testee*, satu jenis tes, dan satu kali testing.

Dengan menggunakan pendekatan *single test-single trial*, maka tinggi rendahnya reliabilitas tes

hasil belajar bentuk obyektif dapat diketahui dengan melihat besar kecilnya koefisien tes, yang dilambangkan dengan  $r_{11}$  atau  $r_{tt}$  (koefisien reliabilitas tes secara total). Adapun untuk mencari atau menghitung  $r_{11}$  atau  $r_{tt}$  dapat digunakan lima jenis formula, yaitu: (1) *formula spearman-brown*, (2) *formula flanagan*, (3) *formula rulon*, (4) *formula kuder-richard-son*, dan (5) *formula C. Hoyt*.

Menggunakan formula *spearman brown*, formula *flanagan* dan formula *rulon*, maka penentuan reliabilitas tes hasil belajar bentuk obyektif dilakukan dengan jalan “membelah dua” tes, karena itu sering dikatakan bahwa ketiga jenis formula itu menggunakan teknik belah dua (*split-half technique*). Sedangkan dengan menggunakan *formula kuder-richardson* dan *formula C. Hoyt* maka penentuan reliabilitas tes hasil belajar bentuk obyektif tersebut tidak menggunakan teknik belah dua.<sup>62</sup>

- 2) Menggunakan pendekatan *test-retest (single test-double trial method)*.

Pada pendekatan *single test-single triad* - dalam rangka penentuan reliabilitas tes hasil belajar bentuk obyektif didasarkan pada konsistensi dari

---

<sup>62</sup> Anas Sudijono, *Pengantar Evaluasi pendidikan*, hlm. 213-214

“batang tubuh” tes hasil belajar yang bersangkutan, yang terbangun dari kumpulan butir-butir item.

Adapun pada pendekatan *single test-double trial* atau pendekatan test-retest, yang juga sering dikenal dengan istilah pendekatan bentuk ulangan, maka penentuan reliabilitas tes dilakukan dengan menggunakan teknik ulangan, di mana tester hanya menggunakan satu seri tes, tetapi percobaannya dilakukan sebanyak dua kali.

Dalam pelaksanaan pengujian reliabilitas tes dengan menggunakan pendekatan *test retest*, skor-skor hasil tes pertama dikorelasikan dengan skor-skor hasil tes kedua. Jika terdapat korelasi positif yang signifikan antara skor-skor hasil tes pertama dengan skor-skor hasil tes kedua, maka tes hasil belajar itu dapat dinyatakan sebagai tes hasil belajar yang reliabel, sebab antara skor-skor hasil tes pertama dengan skor-skor hasil tes kedua memperlihatkan adanya kejegan atau kestabilan.

Untuk mencari korelasi antara skor-skor hasil tes pertama dengan skor-skor hasil tes kedua, dapat dipergunakan teknik korelasi rank-order (teknik korelasi tata jenjang) dari spearman.<sup>63</sup>

---

<sup>63</sup> Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 267-269

3) Menggunakan pendekatan *alternate form (double test-double trial method)*.

Untuk mengetahui apakah tes hasil belajar telah memiliki reliabilitas yang tinggi atautkah belum, dipergunakan dua buah tes yang diberikan kepada sekelompok subyek tanpa adanya tenggang waktu (dilakukan secara berbareng), dengan ketentuan bahwa kedua tes tersebut harus sejenis, dalam arti sekalipun butir-butir itemnya tidak sama, namun hendaknya butir-butir item itu mengukur hal yang sama, baik dari segi isinya, proses mental yang diukur, derajat kesukaran maupun jumlah butir itemnya. Penentuan reliabilitas tes dengan menggunakan pendekatan *alternate form* ini sering dikenal dengan istilah pendekatan bentuk paralel.

Dalam pelaksanaan pengujian reliabilitas tes dengan menggunakan pendekatan *alternate form* atau bentuk paralel ini, skor-skor yang diperoleh dari kedua seri tes tadi dicari korelasinya. Apabila terdapat korelasi positif yang signifikan maka dapat dikatakan bahwa tes hasil belajar tersebut dapat dikatakan reliabel. Teknik korelasi yang dipergunakan bisa dipilih antara teknik korelasi *product moment* dari

Pearson atau teknik korelasi *rank order* dari spearman (khusus untuk N kurang dari 30).<sup>64</sup>

#### 4. Tingkat Kesukaran

##### a. Pengertian tingkat kesukaran

Analisis tingkat kesukaran soal artinya mengkaji soal-soal tes dari segi kesulitan sehingga dapat diperoleh soal-soal yang termasuk muda, sedang dan sukar.<sup>65</sup> Tingkat kesukaran tes item pada umumnya ditunjukkan dengan persentase siswa yang memperoleh jawaban item benar.<sup>66</sup>

##### b. Menentukan proporsi jumlah soal kategori mudah, sedang dan sukar.

Soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang siswa untuk mempertinggi usaha memecahkannya. Sebaliknya soal yang terlalu sukar akan menyebabkan siswa menjadi putus asa dan tidak mempunyai semangat untuk mencoba lagi karena di luar jangkauannya.<sup>67</sup>

---

<sup>64</sup> Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 272-273

<sup>65</sup> Nana Sudjana, *Penilaian Hasil Proses Belajar mengajar*, hlm. 135

<sup>66</sup> Sukardi, *Evaluasi Pendidikan: Prinsip dan Operasionalnya*, hlm.

<sup>67</sup> Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, hlm. 207

Ada beberapa dasar pertimbangan dalam menentukan proporsi jumlah soal kategori mudah, sedang dan sukar. Pertimbangan pertama adalah keseimbangan, yakni jumlah soal sama untuk ketiga kategori tersebut. Artinya soal mudah, sedang dan sukar jumlahnya seimbang. Pertimbangan kedua proporsi jumlah soal untuk ketiga kategori tersebut didasarkan atas kurva normal. Artinya, sebagian besar soal berada dalam kategori sedang, sebagian lagi termasuk ke dalam kategori mudah dan sukar dengan proporsi yang seimbang.

Perbandingan antara soal mudah-sedang-sukar bisa dibuat 3-4-3. Artinya, 30% soal kategori mudah, 40% soal kategori sedang, dan 30% lagi soal kategori sukar.<sup>68</sup>

## 5. Daya Pembeda

### a. Pengertian daya pembeda

Daya pembeda soal adalah kemampuan suatu soal untuk membedakan antara peserta didik yang pandai (menguasai materi) dengan peserta didik yang kurang pandai (kurang atau tidak menguasai materi).<sup>69</sup>

Analisis daya pembeda mengkaji butir-butir soal dengan tujuan untuk mengetahui kesanggupan soal dalam membedakan siswa yang tergolong mampu (tinggi

---

<sup>68</sup> Nana Sudjana, *Penilaian Hasil Proses Belajar Mengajar*, hlm. 135-136

<sup>69</sup> Zainal Arifin, *Evaluasi Pembelajaran: Prinsip Teknik Prosedur*, hlm. 133

prestasinya) dengan siswa yang tergolong kurang atau lemah prestasinya. Artinya, bila soal tersebut diberikan kepada anak yang mampu, hasilnya menunjukkan prestasi yang tinggi, dan bila diberikan kepada siswa yang lemah, hasilnya rendah.<sup>70</sup>

b. Langkah-langkah Menghitung Daya Pembeda

Daya pembeda item itu dapat diketahui melalui atau dengan melihat besar kecilnya angka indeks diskriminasi D (d besar). Daya pembeda pada dasarnya dihitung atas dasar pembagian *testee* ke dalam dua kelompok, yaitu kelompok pandai atau kelompok atas (*upper group*) dan kelompok bodoh atau kelompok bawah (*lower group*).<sup>71</sup>

Adapun cara menentukan dua kelompok itu bisa bervariasi, misalnya dapat menggunakan median sehingga pembagian menjadi dua kelompok itu terdiri atas 50% *testee* kelompok atas dan 50% *testee* kelompok bawah, dapat juga dengan hanya mengambil 20% dari *testee* yang termasuk dalam kelompok atas dan 20% lainnya diambil dari *testee* yang termasuk dalam kelompok bawah, dapat juga menggunakan angka persentase lainnya. Namun pada umumnya para pakar di bidang evaluasi pendidikan lebih banyak menggunakan persentase sebesar 27% dari *testee*

---

<sup>70</sup> Nana Sudjana, *Penilaian Hasil Proses Belajar Mengajar*, hlm. 141

<sup>71</sup> Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, hlm. 211

yang termasuk dalam kelompok atas dan 27% lainnya diambilkan dari *testee* yang termasuk dalam kelompok bawah. Hal ini disebabkan karena berdasarkan bukti-bukti empirik pengambilan subyek sebanyak 27% *testee* kelompok atas dan 27% *testee* kelompok bawah itu telah menunjukkan kesensitifannya, atau dengan kata lain cukup dapat diandalkan.<sup>72</sup>

## 6. Efektifitas Fungsi Distraktor

### a. Pengertian distraktor

Pengecoh (distraktor) dikenal dengan istilah penyesat atau penggoda adalah pilihan jawaban yang bukan merupakan kunci jawaban. Pengecoh diadakan untuk menyesatkan siswa agar tidak memilih kunci jawaban. Pengecoh menggoda siswa yang kurang begitu memahami materi pelajaran untuk memilihnya. Agar dapat melakukan fungsinya untuk mengecoh maka pengecoh harus dibuat semirip mungkin dengan kunci jawaban.<sup>73</sup>

Pengecoh yang tidak dipilih sama sekali oleh *testee* berarti bahwa pengecoh itu jelek, terlalu menyolok menyesatkan. Sebaliknya sebuah distraktor (pengecoh) dapat dikatakan berfungsi dengan baik apabila distraktor tersebut mempunyai daya tarik yang besar bagi pengikut-

---

<sup>72</sup> Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 387

<sup>73</sup> Purwanto, *Evaluasi Hasil Belajar*, hlm. 108

pengikut tes yang kurang memahami konsep atau kurang menguasai bahan. Suatu distraktor dapat dikatakan berfungsi baik jika paling sedikit dipilih oleh 5 % pengikut tes.<sup>74</sup>

b. Tujuan pemakaian distraktor

Tujuan pemakaian distraktor ini adalah mengecohkan mereka yang kurang mampu (tidak tahu) untuk dapat dibedakan dengan yang mampu. Oleh karena itu distraktor yang baik adalah yang dapat dihindari oleh anak-anak yang kurang pandai dan terpilih oleh anak-anak yang kurang pandai.<sup>75</sup>

7. Mata Pelajaran Al-Qur'an Hadits

a. Pengertian mata pelajaran Al-Qur'an

Mata pelajaran Al-Qur'an-Hadis MTs ini merupakan kelanjutan dan kesinambungan dengan mata pelajaran Al-Qur'an-Hadis pada jenjang MI atau SD, terutama pada penekanan kemampuan membaca al-Qur'an-hadis yang baik dan benar yang didasarkan pada ilmu tajwid dan *makhraj al-huruuf*. Di samping itu, ditambah dengan pemahaman secara sederhana terhadap surat-surat pendek, dan mengaitkannya dengan kehidupan sehari-hari, sebagai persiapan untuk belajar sosialisasi

---

<sup>74</sup> Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, hlm. 220

<sup>75</sup> M. Chabib Thoha, *Teknik Evaluasi Pendidikan*, hlm. 149

dengan masyarakat dan melanjutkan pendidikan pada jenjang berikutnya.

b. Tujuan mata pelajaran Al-Qur'an-Hadis

Tujuan mata pelajaran Al-Qur'an-Hadis adalah:

- 1) Meningkatkan kecintaan siswa terhadap al-Qur'an dan hadis.
- 2) Membekali siswa dengan dalil-dalil yang terdapat dalam al-Qur'an dan hadis sebagai pedoman dalam menyikapi dan menghadapi kehidupan.
- 3) Meningkatkan kekhusyukan siswa dalam beribadah terlebih salat, dengan menerapkan hukum bacaan tajwid serta isi kandungan surat/ayat dalam surat-surat pendek yang mereka baca.

c. Ruang lingkup Mata Pelajaran Al-Qur'an Hadits

Ruang lingkup mata pelajaran Al-Qur'an-Hadis di Madrasah Tsanawiyah meliputi:

- 1) Membaca dan menulis yang merupakan unsur penerapan ilmu tajwid.
- 2) Menerjemahkan makna (tafsiran) yang merupakan pemahaman, interpretasi ayat, dan hadis dalam memperkaya khazanah intelektual.
- 3) Menerapkan isi kandungan ayat/hadis yang merupakan unsur pengamalan nyata dalam kehidupan sehari-hari.

d. Standar Kompetensi dan Kompetensi Dasar Mata Pelajaran Al-Qur'an Hadits.<sup>76</sup>

<b>Standar Kompetensi</b>	<b>Kompetensi Dasar</b>
1. Membaca al-Qur'an surat pendek pilihan	1.1. Menerapkan hukum bacaan <i>Qalqalah</i> , <i>tafkhim</i> , dan <i>mad 'aridh lissukun</i> dalam al-Qur'an 1.2. Menerapkan hukum bacaan <i>nun mati</i> , dan <i>mim mati</i> dalam al-Qur'an
2. Menerapkan al-Qur'an surat-surat pendek pilihan dalam kehidupan sehari-hari tentang ketentuan rezeki dari Allah	2.1. Memahami isi kandungan QS <i>al-Quraisy</i> dan <i>al-Insyiraah</i> tentang ketentuan rezeki dari Allah 2.2. Memahami keterkaitan isi kandungan QS <i>al-Quraisy</i> dan <i>al-Insyiraah</i> tentang ketentuan rezeki dari Allah dalam kehidupan. 2.3. Menerapkan isi kandungan QS <i>al-Quraisy</i> dan <i>al-Insyiraah</i> tentang ketentuan rezeki dari Allah dalam kehidupan
3. Menerapkan al-Qur'an surat-surat pendek pilihan dalam kehidupan sehari-hari tentang kepedulian sosial	3.1. Memahami isi kandungan QS <i>al-Kautsar</i> dan <i>al-Maa'un</i> tentang kepedulian sosial. 3.2. Memahami keterkaitan isi kandungan QS <i>al-Kautsar</i> dan <i>al-Maa'un</i> tentang kepedulian sosial dalam fenomena kehidupan
4. Memahami hadis tentang tolong-menolong dan mencintai anak yatim	4.1. Menulis hadis tentang tolong-menolong dan mencintai anak yatim.

<sup>76</sup> Permenag RI Nomor 2 Tahun 2008

Standar Kompetensi	Kompetensi Dasar
	4.2. menerjemahkan makna hadis tentang tolong menolong dan mencintai anak yatim 4.3. Menghafal hadis tentang tolong-menolong dan mencintai anak yatim. 4.4. Menjelaskan keterkaitan isi kandungan hadis dalam perilaku tolong menolong dan mencintai anak yatim dalam fenomena kehidupan dan akibatnya

## B. Kajian Pustaka

Kajian pustaka pada dasarnya digunakan untuk memperoleh suatu informasi tentang teori yang ada kaitannya dengan judul penelitian dan digunakan untuk memperoleh landasan teori ilmiah. Dalam kajian pustaka ini peneliti menelaah beberapa karya ilmiah antara lain:

Penelitian Nur Barri, NIM 07311156 Mahasiswa IAIN Walisongo Semarang Fakultas Tarbiyah tahun 2011, dengan judul “*Analisis Tes Multiple Choice Buatan KKMTs Mata Pelajaran Sejarah Kebudayaan Islam Kelas VII Semester II di M.Ts. Negeri Gondang Sragen tahun 2010/2011*”. Penelitian ini bertujuan mengetahui validitas butir tes mata pelajaran sejarah kebudayaan Islam buatan KKMTs kelas VII semester II memiliki validitas sedang yaitu 57,5%. Kedua memiliki reliabilitas tinggi dengan koefisien korelasi  $r_{11} = 0,797$ . Ketiga dilihat dari tingkat kesukaran terdapat 67.5% termasuk dalam kategori sedang/ cukup,

keempat dilihat dari daya pembeda menunjukkan 42.5%, fungsi distraktor termasuk memiliki distraktor yang baik, yaitu sebesar 76%.<sup>77</sup>

Penelitian Saudara Subakir, NIM 3100218 Mahasiswa IAIN Walisongo Semarang Fakultas Tarbiyah tahun 2007, dengan judul “*Analisis Butir Tes Multiple Choice Mata Pelajaran Pendidikan Agama Islam Semester 1 Siswa Kelas X SMA Negeri 7 Semarang Tahun 2006/2007*”. Hasil penelitian ini menunjukkan tes pada mata pelajaran Pendidikan Agama Islam semester 1 kelas X ini termasuk dalam kategori tes yang memiliki validitas sedang yaitu 58 %, kedua memiliki reliabilitas yang baik dengan koefisien korelasi  $r_{11} = 0,709$ , ketiga dilihat dari tingkat kesukaran terdapat yaitu 50 % termasuk dalam kategori sedang/cukup, keempat dilihat dari daya pembeda menunjukkan kategori sedang 36 %, fungsi distraktor termasuk dalam kategori cukup atau sedang 46,5 %.<sup>78</sup>

Muahadah (NIM 3101217) Mahasiswa IAIN Walisongo Semarang Fakultas Tarbiyah tahun 2006, *Kualitas Instrumen Tes Pendidikan Agama Islam Buatan Guru MGMP Pendidikan Agama*

---

<sup>77</sup> Nur Barri, “Analisis Tes Multiple Choice Buatan KKMTs Mata Pelajaran Sejarah Kebudayaan Islam Kelas VII Semester II di MTs Negeri Gondang Sragen”, *Skripsi*, (Semarang: Fakultas Tarbiyah IAIN Walisongo Semarang, 2010)

<sup>78</sup> Subakir, “Analisis Butir Tes Multiple Choice Mata Pelajaran Pendidikan Agama Islam Semester 1 Siswa Kelas X SMA Negeri 7 Semarang Tahun 2006/2007”, *Skripsi*, (Semarang: Fakultas Tarbiyah IAIN Walisongo Semarang, 2007).

*Islam Di SMPN 1 Demak* Pertama memiliki validitas sedang yaitu 76,7%, kedua memiliki reliabilitas cukup atau sedang dengan koefisien reliabilitas ( $r_{11}$ ) sebesar 0,691, ketiga dilihat dari tingkat kesukaran terdapat 80% termasuk dalam kategori mudah, keempat memiliki daya pembeda sedang/cukup dengan persentase 63,6%, kelima dari option distraktor yang dipasang sudah berfungsi dengan baik sudah cukup efektif yaitu 50%.<sup>79</sup>

Meskipun hampir memiliki kesamaan dengan karya peneliti sebelumnya, yakni membahas tentang validitas dan reliabilitas tes, namun secara prinsipil memiliki perbedaan, yakni pada fokus atau obyek penelitian. Fokus pada penelitian ini adalah validitas dan reliabilitas tes buatan MGMP L.P. Ma'arif kabupaten Kendal pada mata pelajaran Al-Qur'an Hadits semester Gasal Tahun Pelajaran 2013/2014. Jadi penelitian-penelitian yang ada tersebut hanya dijadikan gambaran dan referensi saja oleh peneliti.

---

<sup>79</sup> Muahadah, "Kualitas Instrumen Tes Pendidikan Agama Islam Buatan Guru MGMP Pendidikan Agama Islam di SMP N 1 Demak (NIM 3101217) Mahasiswa IAIN Walisongo Semarang Fakultas Tarbiyah tahun 2006", Skripsi, (Semarang: Fakultas Tarbiyah IAIN Walisongo Semarang, 2006).