

## **BAB II**

### **TINJAUAN TEORI**

#### **A. *Assessment***

##### **1. Pengertian**

*Assessment* merupakan proses pemberian nilai terhadap hasil belajar yang dilakukan pada mahasiswa dengan kriteria tertentu (Azhari, 2016). Penilaian (*assessment*) proses mengumpulkan informasi untuk mengetahui apakah tujuan pembelajaran yang sudah direncanakan sudah tercapai (Nuriyah, 2014).

Berdasarkan pengertian diatas, penilaian (*assessment*) merupakan proses mengumpulkan informasi dengan pemberian nilai terhadap hasil belajar dan rencana pembelajaran yang direncanakan sejak awal, untuk mengetahui apakah tujuan pembelajaran sudah tercapai atau belum.

##### **2. Komponen penilaian**

Dalam melakukan penilaian terhadap hasil belajar mencakup 3 komponen yaitu:

###### **a. Kognitif**

kognitif yaitu kemampuan seseorang dalam memahami, menerapkan, dan melakukan analisis kembali terhadap apapun yang telah dipelajari dengan pengembangan kemampuan intelektual (Elis, 2011). Ranah pembahasan kognitif berhubungan dengan pemahaman dan pengetahuan tentang suatu hal. Aspek kognitif dapat dinilai dengan mengukur pencapaian hasil belajar dari segi intelektualitas atau

pengetahuan, baik itu kemampuan menggali informasi dan kemampuan mengolah kembali suatu hal yang telah didapatkan (Heribertus, 2009).

b. Afektif

Penilaian afektif mencakup penerimaan, penanggapan, perhitungan, dan pengelolaan. Aspek afektif memuat kehendak dan dorongan yang menjadi unsur pembentukan sikap hidup (Heribertus, 2009).

c. Psikomotor

Mencakup dari gerakan fisik dan komunikasi. Ranah psikomotor adalah pembahasan dan penilaian yang berhubungan dengan keterampilan atau aktivitas fisik (Wicaksono, 2016).

3. Tujuan

Penilaian terhadap hasil belajar mahasiswa merupakan bagian dari proses Pendidikan (Azhari, 2016). Tujuan dari penilaian tersebut meliputi:

- a. Untuk melihat apakah tujuan pembelajaran telah dicapai.
- b. Memberikan penilaian terhadap kompetensi mahasiswa.
- c. Sebagai dasar untuk melakukan evaluasi dari program pembelajaran.

4. Metode penilaian

Penilaian hasil belajar berdasarkan taksonomi dibagi menjadi area kognitif, afektif, dan psikomotor.

a. Kognitif

Hasil penilaian kognitif dipusatkan pada pengetahuan dan kemampuan untuk mengingat sampai kemampuan memecahkan suatu masalah. Area kognitif adalah area yang melibatkan kegiatan intelektual (otak) (Wicaksono, 2016). Proses belajar yang menggunakan pengetahuan dimulai dari ketika semua aktivitas dari luar diterima oleh sensori, kemudian disimpan dan diolah dalam otak menjadi informasi, hingga penggunaan kembali informasi tersebut ketika diperlukan untuk penyelesaian masalah (Elis, 2011).

*Bloom's Taxonomy* dalam Wicaksono (2016) membagi area kognitif menjadi 6 tingkatan, yaitu:

- 1) Tingkatan paling dasar yaitu pengetahuan (*Knowledge*), yang didalamnya mencakup kemampuan untuk mengingat dan menghafal.
- 2) Setelah pengetahuan, tingkatan berikutnya adalah pemahaman (*Comprehension*), yang meliputi kemampuan untuk menerjemahkan bahasa, menginterpretasikan bahasa, dan menyimpulkan suatu masalah atau pemikiran.
- 3) Setelah didapatkan pemahaman maka pengetahuan dapat diterapkan (*Application*), meliputi kemampuan untuk memecahkan masalah dengan menerapkan pengetahuan.

- 4) Pada tingkatan berikutnya adalah ranah *Analysis*, meliputi kemampuan untuk menganalisa, memecahkan masalah dan menentukan hubungan diantaranya dengan jelas.
- 5) Setelah analisis dilakukan, maka domain berikutnya adalah *Synthesis*, yaitu kemampuan seseorang untuk menyusun bagian yang terpisah menjadi satu kesatuan.
- 6) Tingkatan yang terakhir yaitu *Evaluation*, yaitu kemampuan untuk menilai dan mempertimbangkan berdasarkan norma internal dan eksternal serta membuat keputusan yang bermakna.

Wicaksono (2016) mengatakan bahwa penilaian ranah kognitif untuk mendapatkan hasilnya memerlukan beberapa instrumen seperti:

1) *True False item*

*True False Item* adalah instrumen untuk mengetahui *skill* kognitif siswa pada tiap jenjang pendidikan. Tes berupa pertanyaan sederhana dan singkat. Persoalan yang siswa harus pecahkan adalah kemungkinan jawaban benar (*true*) atau salah (*false*)

2) *Extended Matching item (EMI)*

EMI tersusun dari satu tema, pilihan jawaban yang memungkinkan, dasar untuk menghubungkan, dan pertanyaan yang berarah (*Lead – in*), serta suatu skenario. EMI digunakan untuk kemampuan mengingat saja

### 3) *Multiple Choice Question (MCQ)*

MCQ merupakan metode evaluasi yang paling sering digunakan. MCQ dapat mengevaluasi bagian pengetahuan secara cepat dan bisa digunakan untuk kelompok mahasiswa yang banyak. Penggunaan bentuk soal yang terlalu rumit tidak dianjurkan karena dapat membingungkan mahasiswa.

#### b. Afektif

Hasil difokuskan pada perilaku dalam menggunakan sesuatu, terhadap teman sejawat, emosional, serta perasaan. Seperti penerimaan, pemberian respon/partisipasi, penghargaan terhadap nilai, pengorganisasian (menghubungkan nilai-nilai yang telah dipegang dan mengintegrasikan nilai-nilai tersebut dalam kehidupannya), dan pengalaman (Qadar *et al*, 2015).

#### c. Psikomotor

Hasil yang difokuskan pada *skill* atau keterampilan. Ranah psikomotor berkaitan dengan keterampilan atau kemampuan bertindak, setelah seseorang menerima pengalaman belajar tertentu (Elis, 2011). Seperti *imitation* (menirukan gerak yang telah diamati), *manipulation* (menggunakan konsep yang sama untuk melakukan gerak), *precision* (melakukan gerak dengan teliti dan benar), *articulation* (merangkaikan berbagai gerakan secara berkesinambungan), dan *naturalisation* (melakukan gerakan secara wajar dan efisien).

## **B. *Multiple Choice Question (MCQ)***

### 1. Pengertian

*Multiple choice question (MCQ)* adalah jenis evaluasi yang menggunakan pilihan ganda. Biasanya MCQ digunakan untuk menilai kemampuan kognitif seseorang. Instrumen evaluasi yang paling sering digunakan adalah MCQ. MCQ dapat mengevaluasi bagian pengetahuan secara cepat dan bisa digunakan untuk kelompok mahasiswa yang banyak. MCQ memiliki kelemahan yaitu tidak dapat menilai area perilaku dan area keterampilan. MCQ berpotensi untuk mengukur pemahaman dan pengetahuan seseorang terhadap sesuatu hal yang telah dipelajari. Penggunaan bentuk soal yang terlalu rumit tidak dianjurkan karena dapat membingungkan mahasiswa (Shafira, 2015).

### 2. Analisis soal MCQ

MCQ banyak digunakan di institusi pendidikan untuk menguji hasil belajar siswa. MCQ biasanya terdiri dari sebuah pertanyaan atau pernyataan yang tidak lengkap yang disebut sebagai *stem*, dan serangkaian dua atau lebih pilihan yang salah satunya adalah jawaban paling tepat untuk jawaban tersebut. Tugas pelajar pada tes MCQ adalah memilih salah satu jawaban yang paling sesuai dengan soal (Quaigrain and Arhin, 2017). Jawaban yang paling tepat disebut sebagai kunci dan jawaban lainnya disebut sebagai distraktor, hanya ada 1 jawaban yang harusnya paling benar dan distraktor harusnya tidak tepat. Salah satu kesulitan dari pembuatan soal MCQ adalah membuat distraktor yang bagus, yang tidak

terlalu mengecoh mahasiswa, dan juga tidak terlalu jelas. Salah satu faktor yang dapat menunjukkan kualitas soal MCQ adalah adanya distraktor yang bagus (*Rao et al, 2017*).

Salah satu perhatian utama dalam pembuatan soal MCQ untuk ujian adalah reliabilitas nilai tes. Biasanya, MCQ yang baik adalah siswa yang mengalami proses dapat melakukan analisis item yang ketat. Analisis item adalah proses mengumpulkan, meringkas, dan menggunakan informasi dari tanggapan siswa untuk menilai kualitas item tes. Analisis item dapat digunakan untuk mengamati karakteristik dari item tertentu dan dapat digunakan untuk memastikan bahwa item tersebut memiliki standar yang sesuai untuk dimasukkan dalam pengujian, atau jika item tersebut memerlukan perbaikan (*Quaigrain and Arhin, 2017*).

Analisis item dapat digunakan untuk mengamati karakteristik item, dan untuk meningkatkan kualitas pengujian (*Rao et al, 2017*). Menurut *Quaigrain and Arhin (2017)* revisi item memungkinkan kita mengidentifikasi item terlalu sulit atau terlalu mudah, item tidak dapat membedakan antara siswa yang telah mempelajari konten dan mereka yang tidak memilikinya, atau pertanyaan yang memiliki distraktor yang tidak masuk akal dengan bijak, para pengajar dapat menghapus soal-soal non-diskriminasi dari kumpulan soal atau mengubah soal atau memodifikasi instruksi untuk memperbaiki kesalahpahaman tentang konten atau menyesuaikan dengan cara mereka mengajar.

Kualitas soal dapat dinilai dari tingkat kesulitannya dan kemampuan soal dalam membedakan antara mahasiswa yang benar-benar memahami pelajaran dengan mahasiswa yang memahami secara belum sempurna (*faulty grasp*) yang disebut dengan tingkat diskriminasi dan jumlah *distractor effectiveness*. Penentuan tingkat kesulitan dan tingkat diskriminasi ini dilakukan dalam Item Analisis soal. Item Analisis sebaiknya dilakukan secara berkesinambungan (Taib and Yusoff, 2014).

Untuk analisis soal MCQ terdiri dari 3 tingkatan, yaitu:

a. *Item difficulty*

Item Indeks kesulitan dihitung sebagai persentase dari jumlah total jawaban yang benar terhadap item tes. Hal ini dihitung dengan menggunakan rumus:

$$p = \frac{R}{T}$$

Dimana p adalah indeks kesulitan item, R adalah jumlah jawaban yang benar, dan T adalah jumlah total soal (Quaigrain and Arhin, 2017).

Kesulitan indeks (*p-value*), juga disebut kemudahan indeks, menggambarkan persentase siswa yang menjawab dengan benar item tersebut. Ini berkisar antara 0 - 100%. Semakin tinggi persentasinya, semakin mudah itemnya. Kisaran kesulitan yang disarankan adalah dari 30 - 70%. Item yang memiliki nilai p di bawah 30% dan di atas 70% masing-masing dianggap sebagai item yang sulit dan mudah. Item yang sangat mudah biasanya harus ditempatkan pada awal tes sebagai



pertanyaan 'pemanasan' atau dihilangkan sama sekali. Item yang sulit harus ditinjau ulang untuk kemungkinan bahasa yang membingungkan, area kontroversi, atau bahkan kunci yang salah. Pencantuman item yang sangat sulit dalam ujian tergantung pada target guru, yang mungkin ingin memasukkannya untuk mengidentifikasi siapa yang mendapat nilai terbanyak (Hingorjo and Jaleel, 2012).

*b. Item discrimination*

Item *discrimination index* adalah korelasi biserial titik antara mendapatkan item yang benar dan nilai total pada semua item lainnya. Kemudian, jumlah siswa di atas 27% yang memperoleh tanggapan yang benar dan yang terendah 27% yang memperoleh tanggapan yang benar dihitung. Discrimination index dihitung menggunakan rumus :

$$DI = \frac{UG-LG}{n}$$

Dimana UG adalah jumlah siswa di kelompok atas yang mendapat item yang benar dan LG adalah jumlah siswa di kelompok bawah yang mendapat item yang benar dan n adalah jumlah orang yang terbesar dari kedua kelompok. Semakin tinggi DI, semakin baik item tes membedakan antara siswa dengan nilai tes yang lebih tinggi dan skor tes yang lebih rendah (Quaigrain and Arhin, 2017).

Indeks Diskriminasi (DI), juga disebut korelasi titik biserial (PBS), menggambarkan kemampuan item untuk membedakan antara pencetak skor tinggi dan rendah, ini berkisar antara -1.00 dan +1.00. Diharapkan siswa yang berkinerja tinggi memilih jawaban yang benar

untuk setiap item lebih sering daripada siswa berprestasi rendah. Jika ini benar, penilaian dikatakan memiliki DI positif (antara 0,00 dan +1.00), menunjukkan bahwa siswa yang mendapat nilai total tinggi, memilih jawaban yang benar untuk item tertentu lebih sering daripada siswa yang memiliki keseluruhan rendah skor. Jika, bagaimanapun, siswa berkinerja rendah mendapatkan barang tertentu dengan lebih tepat daripada pencetak skor tinggi, maka item tersebut memiliki nilai DI negatif (antara -1.00 dan 0.00) (Hingorjo and Jaleel, 2012).

Berdasarkan pedoman Ebel (1979) dalam Quaigrain and Arhin (2017) tentang item analisis item teori uji klasik dikategorikan dalam indeks diskriminatif mereka. Sebagai aturan praktis:

- 1) Jika  $DI \geq 0,40$ , maka soal tersebut berfungsi dengan memuaskan.
- 2) Jika  $0,30 \leq DI \leq 0,39$ , diperlukan revisi sedikit atau tidak ada.
- 3) Jika  $0,20 \leq DI \leq 0,29$ , maka item tersebut adalah marjinal dan revisi kebutuhan.
- 4) Jika  $DI \leq 0,19$ , maka soal tersebut harus dieliminasi atau direvisi sepenuhnya.

c. *Distractor effectiveness of alternatives*

Item berisi *stem* dan empat pilihan termasuk satu pilihan yang benar (kunci) dan tiga kesalahan (distraktor). *Nonfunctional distractor (NFD)* pada item adalah pilihannya, selain kunci jawaban yang dipilih oleh <5% siswa dan distraktor fungsional atau efektif adalah pilihan yang dipilih oleh 5% atau lebih siswa. Atas dasar NFS pada suatu soal,

*distractor effectiveness (DE)* berkisar antara 0% sampai 100%. Jika item berisi tiga atau dua atau satu NFD, maka DE masing-masing adalah 0.33,3%, 66.6%, dan 100% (Rao et al, 2017).

Analisis distraktor adalah bagian penting lainnya dari analisis item. Para distraktor adalah komponen penting dari sebuah item, karena menunjukkan hubungan antara total skor tes dan distraktor yang dipilih oleh siswa. Kinerja siswa bergantung pada bagaimana distraktor dirancang. Efisiensi distractor adalah salah satu alat yang memberi tahu apakah benda itu dibangun dengan baik atau gagal untuk melakukan tujuannya. Setiap distraktor yang telah dipilih oleh kurang dari 5% siswa dianggap sebagai distractor yang tidak berfungsi (NF-D).<sup>7</sup> Idealnya, siswa dengan skor rendah, yang belum menguasai subjek, harus lebih memilih para distraktor. , sedangkan pencetak skor tinggi harus membuangnya lebih sering saat memilih opsi yang benar. Dengan menganalisis distraktor, menjadi lebih mudah untuk mengidentifikasi kesalahan mereka, sehingga bisa direvisi, diganti, atau dihapus (Hingorjo and Jaleel, 2012).

### **C. Hasil MCQ**

Hasil MCQ merupakan hasil yang telah dicapai mahasiswa dalam bentuk angka-angka atau skor setelah diberikan tes MCQ. Nilai yang diperoleh mahasiswa menjadi acuan untuk melihat pemahaman dan penguasaan mahasiswa dalam menerima materi pembelajaran.

Data hasil MCQ berupa data ordinal. Data ordinal merupakan data himpunan yang berupa tingkatan, pangkat, dan jabatan. Pemberian skor yang sering digunakan untuk mempermudah dalam mengkategorikan peringkat dari hasil MCQ yang telah didapatkan dituliskan dalam presentase. Pembagian kategori penilaian untuk hasil ujian menurut Nursalam (2009) yaitu:

1. Baik

Apabila jumlah nilai dari jawaban yang benar yaitu 75-100.

2. Cukup

Apabila jumlah nilai dari jawaban yang benar yaitu 60-74.

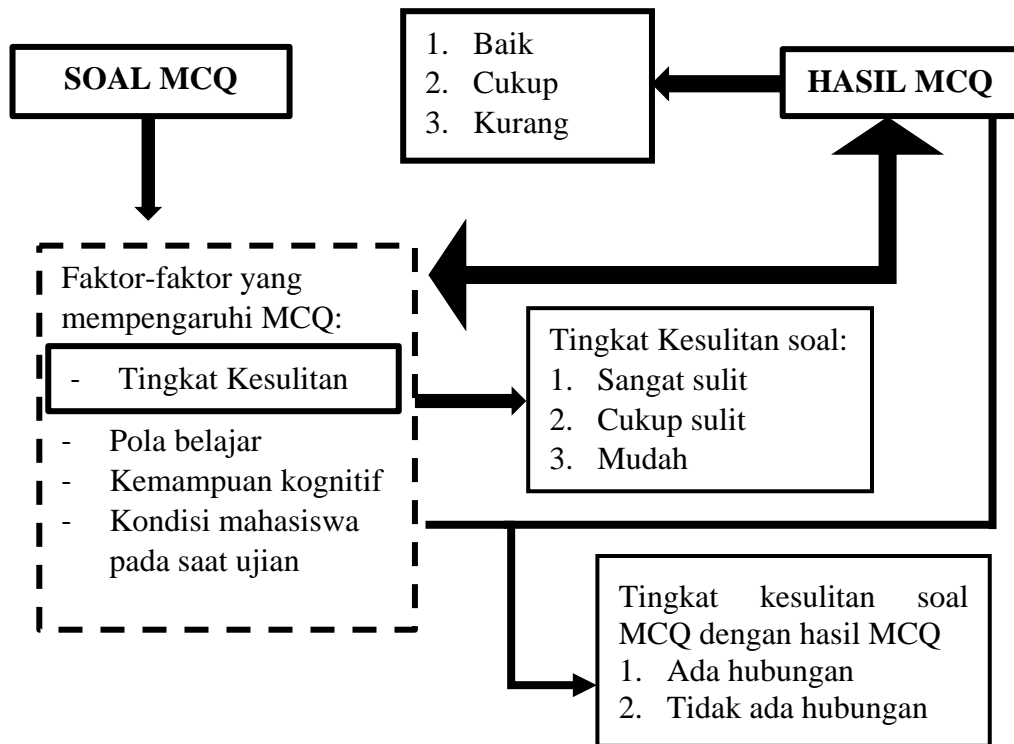
3. Kurang

Apabila jumlah nilai dari jawaban yang benar yaitu 41-59.

4. Sangat kurang

Apabila jumlah nilai dari jawaban yang benar yaitu  $\leq 40$ .

#### D. Kerangka Konsep



Keterangan:

1. Diteliti =

2. Tidak diteliti =

$H_0$  = Terdapat hubungan antara tingkat kesulitan soal MCQ dengan hasil MCQ mahasiswa PSIK FKIK UMY.

$H_1$  = tidak terdapat hubungan antara tingkat kesulitan soal MCQ dengan hasil MCQ mahasiswa PSIK FKIK UMY.