

## Ketepatan Kontruksi Butir Pilihan Ganda Bahasa Arab

Muhibbatul Laili

Universitas Pendidikan Indonesia, Bandung, Indonesia

E-mail: [zahiroilailimuhibbah@gmail.com](mailto:zahiroilailimuhibbah@gmail.com)

### Abstract:

*The study of construct validity on summative test items in multiple-choice Arabic aims to improve the test items' quality to make them valid in the construction aspect. The study was carried out on multiple-choice items based on the modified construction criteria instrument based on the Guttman scale. Items that match the indicator are marked with S and TS if they do not match. Based on a review of 40 items, it was found that the items were not suitable, namely: 25 items with unformulated stem clearly and firmly, two questions that pointed to the correct answer, and one question with pictures that confused. Meanwhile, other items are in accordance with all construct criteria indicators for multiple-choice questions. Based on the research results, the question maker must correct and pay attention to these aspects to produce a valid test kit.*

### Keywords:

*Construct; Multiple choice; Summatif*

### Abstrak

Telaah validitas konstruk terhadap butir soal tes sumatif bahasa Arab pilihan ganda bertujuan untuk meningkatkan kualitas butir tes agar menjadi valid dalam aspek konstruksi. Telaah dilakukan pada butir soal pilihan ganda berdasarkan instrumen kriteria konstruksi yang dimodifikasi berdasarkan skala Guttman. Butir soal yang sesuai dengan indikator diberikan tanda S dan TS jika tidak sesuai. Berdasarkan telaah 40 butir soal, ditemukan butir soal yang kurang sesuai yaitu: 25 butir soal dengan stem yang tidak terumuskan dengan jelas dan tegas, dua pokok soal yang menunjuk kepada jawaban benar, dan satu soal dengan gambar yang memunculkan kebingungan. Sedangkan butir yang lain telah sesuai dengan seluruh indikator kriteria konstruk untuk soal pilihan ganda. Berdasarkan hasil penelitian tersebut maka pembuat soal harus memperbaiki dan memperhatikan aspek-aspek tersebut agar menghasilkan alat tes yang valid.

### Kata Kunci:

Konstruksi; Pilihan ganda; Sumatif

## PENDAHULUAN

Sebagai bentuk profesionalismenya seorang pendidik harus memiliki sebuah kompetensi dalam menilai capaian hasil belajar (Anwar & Fakhruddin, 2016). Penilaian yang dilakukan oleh pendidik mempunyai peran dalam rangka memonitoring kemajuan belajar, hasil belajar, serta menganalisis kebutuhan dalam rangka memperbaiki hasil belajar siswa secara terus-menerus. Penilaian tersebut dibagi menjadi tiga berdasarkan fungsinya, yaitu formatif, sumatif, dan evaluatif (Direktur Jenderal Pendidikan Islam, 2018).

Penilaian formatif (*assessment for learning*) adalah penilaian yang dilaksanakan saat proses belajar sedang dilakukan demi memberikan umpan balik yang berkelanjutan antara siswa dan guru, sehingga keduanya dapat meningkatkan pembelajaran dan pengajaran (Hawe & Dixon, 2016; Spector, dkk., 2016; Furtak, dkk., 2017). Hal ini sangat bertolak belakang dengan penilaian sumatif

(*assessment of learning*) yang dilakukan untuk menunjukkan penguasaan mata pelajaran setelah menyelesaikan proses pembelajaran (Tempelaar, dkk, 2017). Adapun penilaian evaluatif adalah penilaian yang bertujuan untuk melakukan evaluasi terhadap pengelolaan kegiatan belajar mengajar di suatu kelas ataupun satuan pendidikan (Direktur Jenderal Pendidikan Islam, 2018).

Informasi keberhasilan belajar peserta didik dari satuan pendidikan dan kenaikan kelas di Indonesia saat ini ditentukan melalui penilaian sumatif berupa Penilaian Akhir Tahun (PAT). Hal ini selaras dengan pendapat yang mengatakan bahwa penilaian sumatif dilakukan di akhir pembelajaran (Suharismi, 2011; Syihabuddin, 2018) untuk menentukan level kemampuan peserta didik dalam materi pembelajaran yang diajarkan secara keseluruhan serta menentukan naik kelas atau tidaknya (Am, 2019). Selain itu *National Researcher Council* (Dewan Riset Nasional) Washington menyatakan bahwa penilaian ini mempunyai tujuan untuk mengetahui apa saja yang telah ditangkap oleh siswa selama belajar, mengetahui kualitas pembelajaran, dan menilai hasil yang berdasarkan standar-standar tertentu (Dixon & Worrell, 2016; Broadbent, dkk., 2017). Direktur Jenderal Pendidikan Islam (Dirjen Pendis) Nomor 5162 Tahun 2018 yang membahas terkait petunjuk teknis dalam penilaian hasil belajar pada Madrasah Tsanawiyah, memberikan definisi lebih spesifik tentang penilaian sumatif yang diselenggarakan di lingkungan Madrasah, menurutnya penilaian sumatif adalah adalah jenis penilaian yang dengannya dapat terkumpul informasi-informasi tentang yang dilakukan dalam jangka waktu tertentu atau di akhir sebuah unit pelajaran (Direktur Jenderal Pendidikan Islam, 2018).

Indonesia telah menerapkan penilaian dengan dua teknik, yaitu penilaian tes dan penilaian non tes. Untuk melakukan penilaian sumatif, pendidik atau satuan pendidikan dapat melakukannya melalui pemberian tes, sebagaimana yang diamanatkan oleh negara (Dixon & Worrell, 2016). Tes adalah teknik yang paling sering dimanfaatkan dalam kegiatan pengukuran atau penilaian (Arifin, 2012). Teknik tes dapat dilakukan dengan berbagai macam bentuk soal, di antaranya soal pilihan ganda yang disajikan melalui lembar soal sebagai ujian tertulis. Selain disajikan sebagai ujian tertulis, penyajian soal tersebut dapat diselenggarakan dengan media TIK (Hakim & Nugraha, 2016), baik komputer maupun android. Sehingga waktu dapat dihemat, serta lebih efektif dan efisien.

Dalam menghasilkan informasi yang sesuai dengan realita diperlukan instrumen tes yang tepat (Hikmah, 2019). Pilihan ganda lebih banyak digunakan untuk tes di dunia pendidikan, karena

memungkinkan untuk melakukan banyak penilaian, mulai dari penilaian pengetahuan, keterampilan bahkan kompetensi secara efisien dan objektif (Gierl, dkk, 2017). Tes pilihan ganda tersusun atas sebuah pokok soal (*stem*), beberapa pilihan jawaban (*option*) dengan satu kunci jawaban (*key*) yang merupakan jawaban benar dan lainnya sebagai pengecoh (*distracor*), serta tambahan informasi lain yang mendukung (Shin, dkk., 2019).

Tes pilihan ganda mempunyai beberapa keunggulan dan kelebihan (Pusat Penelitian Pendidikan, 2019). Di antara keunggulan yang dimilikinya adalah: (1) mampu mengukur berbagai jenjang sekaligus, (2) penskorannya bisa dilaksanakan dengan efisien serta meliputi ruang lingkup bahan/materi/pokok bahasan yang luas, dan (3) lebih tepat dilaksanakan untuk tes dengan jumlah peserta banyak atau bersifat masal. Akan tetapi, tes ini mempunyai kekurangan berupa: (1) dibutuhkan waktu lama dalam penulisan soal, (2) pembuat soal biasanya merasa kesulitan dalam menghasilkan pilihan jawaban pengecoh homogen serta mempunyai fungsi, dan (3) didapatinya peluang untuk menebak kunci jawaban.

Sebagaimana yang dituangkan dalam Keputusan Direktur Jenderal Pendidikan Islam Nomor 5162 Tahun 2018 tentang Petunjuk Teknis Penilaian Hasil Belajar Pada Madrasah Tsanawiyah (Direktur Jenderal Pendidikan Islam, 2018), dalam proses pengembangan soal setiap pengembang harus memperhatikan rambu-rambu dalam menulis butir soal yang mencakup substansi/materi, konstruksi, dan bahasa (Syihabuddin, 2018). Akan tetapi permasalahan lain muncul ketika guru menyusun instrumen tes, yaitu kelemahan guru dalam proses penyusunannya (Pratama, 2020). Hal ini mengakibatkan instrumen tes tidak valid, sehingga daya ukurnya pun tidak valid.

Dalam pelaksanaannya, MTs. Hasyim Asy'ari Batu merupakan salah satu madrasah yang melaksanakan penilaian sumatif berbasis android dan komputer melalui tes dengan seluruh soal berupa pilihan ganda. Pada tingkat MTs, pada jenjang MTs, pilihan ganda yang disediakan sebanyak empat pilihan, dengan satu kunci (*key*) yang merupakan jawaban yang paling tepat dan lainnya sebagai pengecoh (*distracor*) (Direktur Jenderal Pendidikan Islam, 2018). Selain itu, dalam setiap tes yang disusunnya guru harus selalu mempertimbangkan kualitas butir soal yang akan diteskan kepada peserta didik. Kualitas yang dimaksud adalah validitas isi dan penyusunannya yang sudah memenuhi kriteria kaidah penulisan soal yang ditetapkan. Berdasarkan observasi dan wawancara yang telah dilakukan sebelumnya, guru tidak mempertimbangkan validitas dari tes

yang dibuatnya. Guru membuat soal hanya mempertimbangkan materi yang telah disampaikan dan mengabaikan aspek validitas soal yang dibuatnya.

Dengan demikian setiap pendidik wajib menelaah soal-soal yang dibuatnya agar sesuai dengan ketentuan penyusunan soal agar menghasilkan butir soal yang valid serta dapat menggambarkan informasi tentang peserta didik. Sedangkan kualitas validitas butir soal tes sumatif yang dikembangkan oleh guru bahasa Arab pada di MTs. Hasyim Asy'ari Batu belum ditelaah berdasarkan kriteria penyusunan soal. Sehingga pada penelitian kali ini peneliti melakukan telaah terhadap butir soal yang dikembangkan oleh guru bahasa Arab MTs. Hasyim Asy'ari Batu.

Penelitian terkait analisis butir soal bahasa Arab telah dilakukan sebelumnya oleh Al Fathiyah (2019). Penelitiannya dilakukan pada dokumen soal dan hasil jawaban siswa melalui pendekatan kuantitatif. Metode anates yang digunakannya untuk analisis data dijabarkan berdasarkan tingkat validitas, reliabilitas, kesukaran, daya beda dan pengecoh pada tiap butir. Berdasarkan hasil penelitiannya ditemukan sebanyak 97.15% butir tidak valid, tingkat reliabilitas 0.54 yang menunjukkan arti sedang, lebih dari 50% butir dengan kategori mudah dan sebagian besar soal tidak mempunyai daya beda. Sehingga implikasinya soal tersebut direvisi atau tidak digunakan.

Kualitas butir soal bahasa Arab juga telah dilakukan oleh Fauziyah (2020) dengan sumber data berupa soal HOTS. Fokus penelitian serupa dengan Al Fathiyah. Soal tersebut diketahui memiliki tingkat validitas dan reliabilitas yang tinggi. Akan tetapi 62,5% soal tidak sesuai dengan kaidah penulisan pilihan ganda. Selain itu tingkat kesukaran belum proporsional dengan daya pembeda sedang dan efektifitas pengecoh masuk ke dalam kategori cukup. Namun sayangnya seluruh soal yang di tes merupakan soal LOTS, bukan HOTS seperti yang seharusnya. Berbeda dengan penelitian Al Fathiyah yang menguji validitas berdasarkan daya beda soal dan Fauziyah yang menguji validitas berdasarkan kaidah secara umum, penelitian ini mempunyai fokus pada salah satu aspek, yaitu aspek validitas konstruksi, karena ditemukan kejanggalan dari segi konstruk pada soal ini.

Selanjutnya, penelitian Maulana & Sanusi (2020) yang meneliti kualitas butir-butir dan soal-soal mata pelajaran bahasa Arab yang digunakan pada UAMBD MI tahun ajaran 2017-2018.

Ditinjau dari segi validitas dan reliabilitasnya. Hasil penelitiannya menunjukkan bahwa validitas dan reliabilitas soal bahasa Arab menunjukkan sangat tinggi.

Berbeda dengan penelitian ini, bahwa fokus penelitian ini adalah menganalisis dokumen butir soal yang dilakukan pada penilaian sumatif bahasa Arab melalui tes berbentuk pilihan ganda berbasis komputer dan android di MTs. Hasyim Asy'ari Batu.

## METODE

Penelitian yang bertujuan untuk menganalisis dokumen butir soal ini dilakukan pada penilaian sumatif bahasa Arab yang dilakukan melalui tes berbentuk pilihan ganda berbasis komputer dan android di MTs. Hasyim Asy'ari Batu. Penelitian dilakukan dengan pendekatan deskriptif kualitatif. Peneliti menggambarkan dan mendeskripsikan secara jelas kualitas konstruk dari butir soal tersebut.

Teknik pengumpulan data dilakukan melalui dokumentasi. Dokumen soal penilaian sumatif berupa tes berbentuk pilihan ganda dari guru mata pelajaran bahasa Arab di MTs. Hasyim Asy'ari Batu dikumpulkan demi terlaksananya teknik tersebut.

Sedangkan dalam analisis data, peneliti menggunakan instrumen telaah butir soal yang tercantum dalam buku *Panduan Penilaian Tes Tertulis* yang diterbitkan oleh Pusat Penilaian Pendidikan tahun 2019, dan dimodifikasi menggunakan skala Guttman. Skala Guttman digunakan untuk memandu identifikasi atau definisi konstruk (Antipkina & Ludlow, 2020).

**Tabel 1.**  
Format Instrumen Kajian Butir Soal

Mata Pelajaran : \_\_\_\_\_

Jumlah/Bentuk Soal : \_\_\_\_\_

Isilah dengan menggunakan huruf "S" jika butir soal sesuai dengan indikator aspek yang dinilai, dan "TS" jika butir soal tidak sesuai.

Indikator Konstruk yang Ditelaah	Jumlah Butir	
	Sesuai	Tidak Sesuai
1. Pokok soal dirumuskan secara jelas dan tegas.		
2. Rumusan pokok soal dan pilihan jawaban merupakan pernyataan yang diperlukan saja.		
3. Pokok soal tidak memberi petunjuk ke arah jawaban benar.		

4. Pokok soal tidak mengandung pernyataan yang bersifat negatif ganda.
5. Panjang rumusan pilihan jawaban relatif sama.
6. Pilihan jawaban tidak mengandung pernyataan, "Semua pilihan jawaban di atas salah" atau "Semua pilihan jawaban di atas benar".
7. Pilihan jawaban yang berbentuk angka atau waktu disusun berdasarkan urutan besar kecilnya nilai angka tersebut atau kronologisnya.
8. Gambar, grafik, tabel, diagram, dan sejenisnya yang terdapat pada soal jelas, berfungsi, tidak memunculkan kebingungan, dan mempunyai tingkat keterbacaan tinggi.
9. Butir soal tidak bergantung pada jawaban soal sebelumnya.

## HASIL DAN PEMBAHASAN

Menurut kajian yang telah dilaksanakan menggunakan format telaah butir soal yang tertuang dalam Panduan Penilaian Tertulis 2019, ditemukan data seperti yang ada dalam gambar berikut:

**Tabel 2.**

Hasil Telaah Butir Soal

No	Jumlah Butir	
	Sesuai	Tidak Sesuai
1	15	25
2	40	0
3	38	2
4	40	0
5	40	0
6	40	0
7	40	0
8	39	1
9	40	0

### 1. Pokok soal dirumuskan secara jelas dan tegas

Pada indikator instrumen yang pertama ditemukan 15 soal yang mempunyai rumusan soal yang telah dirumuskan dengan jelas dan tegas, yaitu butir nomor 3, 9, 15, 18, 21, 23, 27, 29,30, 32, 35,36, 37, dan 39. Sedangkan 25 soal lainnya mempunyai rumusan soal yang kurang

jelas dan kurang tegas, yaitu nomor 1, 2, 4, 5, 6, 7, 8, 11, 12, 13, 14, 16, 17, 19, 20, 22, 24, 25, 26, 28, 31, 33, 34, 38, dan 40.

Adapun pokok soal yang belum terumuskan dengan jelas dan tegas, ditemukan tiga macam bentuk. **Pertama, tidak mencantumkan kalimat perintah**, kasus ini terdapat pada butir nomor 1, 2, 4, 5, 6, 7, 8, 11, 12, 13, 14, 16, 20, 22, 24, 25, 26, 28, 31, 33, 34, dan 40. Sebagai gambarannya adalah butir soal nomor 1:

1. يَزْرَعُ النَّبَاتَاتِ وَ الْمَوَادِّ الْغِدَائِيَّةِ . . . . .

“.... dia menanam tumbuh-tumbuhan dan bahan makanan.”

Pada butir tersebut, pokok soal hanya mencantumkan kalimat yang harus dilengkapi, tanpa menyertakan kalimat perintah baik pada baris utama maupun pada setiap butirnya. Untuk melengkapi kata yang rumpang dari kalimat tersebut. Hal ini dapat menyebabkan ketidakjelasan maksud soal tersebut, sehingga menimbulkan pertanyaan bagi peserta didik. Pada pokok soal-soal tersebut, guru dapat menambahkan kalimat yang berisi perintah secara jelas dan tegas untuk melengkapi kalimat rumpang yang tersaji dalam soal-soal tersebut. Contoh pokok soal yang lebih baik dan dapat diterapkan pada butir soal yang mempunyai kasus serupa:

1. يَزْرَعُ النَّبَاتَاتِ وَ الْمَوَادِّ الْغِدَائِيَّةِ . . . . .

إِمْلَأِ الْفَرَاغَ تِلْكَ الْجُمْلَةَ لِتَكُونَ جُمْلَةً مُفِيدَةً!

“.... dia menanam tumbuh-tumbuhan dan bahan makanan.

Lengkapilah kalimat tersebut sehingga menjadi kalimat yang sempurna!”

Dengan memperjelas perintah yang dimaksudkan pada rumusan soal tersebut, siswa tidak perlu lagi menanyakan maksud butir soal tersebut. Peserta didik dapat dengan mudah menangkap perintah yang dimaksud dari pokok soal, sehingga mereka dapat memanfaatkan waktu untuk mengerjakan butir soal yang lain.

**Kedua, kesalahan penggunaan tanda baca pada pokok soal**, sehingga mengurangi ketegasan kalimat. Kasus ini terdapat pada soal nomor 17, 19 dan 38. Kasus nomor 17, terdapat kesalahan pada kalimat tanya. Butir soal berbunyi:

17. ..... أُمِّي مُوظَّفَةٌ, هِيَ تَذْهَبُ كُلَّ يَوْمٍ إِلَى الْإِدَارَةِ . مَا مَعْنَى مُوظَّفَةٌ .....

“Tbuku adalah seorang pegawai, setiap hari ia pergi ke kantor. Apa arti dari kata مُوظَّفَةٌ.....”

Pokok soal berisi sebuah kalimat pernyataan dan sebuah kalimat tanya yang menanyakan makna kata yang bergaris bawah pada kalimat pernyataan sebelumnya. Kalimat interogatif seharusnya diikuti dengan tanda tanya di akhir kalimat. Akan tetapi pada kasus ini, kalimat tersebut mempunyai akhiran tujuh tanda titik. Pada kalimat interogatif tersebut sesungguhnya sudah tampak jelas maksud dari soal, akan tetapi dengan adanya tanda baca titik yang berjumlah tujuh titik pada akhir kalimat serta tidak terdapat tanda tanya pada akhir sebuah kalimat tanya dapat mengurangi ketegasan pokok soal.

Pada kasus butir soal nomor 19, perintah telah dicantumkan dengan jelas, akan tetapi terdapat kesalahan penggunaan tanda baca.

19. Arsitek membangun gedung dan jalan raya.

Terjemahkan ke dalam bahasa Arab? ...

Dalam butir soal tersebut terdapat kalimat perintah yang seharusnya diakhiri dengan tanda seru, akan tetapi yang tercantum adalah tanda tanya yang diikuti tiga titik setelahnya. Sedangkan kasus butir soal nomor 38, terdapat pokok soal yang berisi kalimat perintah, tetapi diakhiri dengan tanda tanya yang diikuti dengan sembilan titik setelahnya, bukan diakhiri dengan tanda seru yang menunjukkan kalimat perintah.

38. ..... أُحِبُّ قِرَاءَةَ الْقُرْآنِ كُلِّ صَبَاحٍ. حَوْلَ الْمَصْدَرِ ( قِرَاءَةٌ ) إِلَى الْفِعْلِ الْمُضَارِعِ؟ .....

“Saya suka membaca Al-Quran setiap pagi. Ubahlah masdar kata ( قِرَاءَةٌ ) menjadi fi’il mudāri’?.....”

Kedua hal tersebut dapat memicu kesalahpahaman siswa terhadap perintah soal yang sesungguhnya. Untuk memperbaikinya guru cukup mengubah tanda tanya menjadi tanda seru tanpa mengikut sertakan tanda titik setelahnya. Contoh soal yang lebih baik:

"Arsitek membangun gedung dan jalan raya.



Terjemahkan kalimat tersebut ke dalam bahasa Arab!"

Dengan bentuk pokok soal tersebut, rumusan pokok soal menjadi jelas dan tegas. Mereka dapat memahami dengan jelas perintah dan maksud dari soal tersebut. Selain itu, tidak menyebabkan munculnya pemahaman ganda yang mengakibatkan peserta didik gamang dalam menentukan pilihan jawaban. Sehingga mereka cukup melewati waktu yang singkat untuk sekedar memahami soal dan menentukan jawaban benarnya.

## 2. Rumusan pokok soal dan pilihan jawaban merupakan pernyataan yang diperlukan saja

Hasil telaah dari seluruh butir soal tersebut, tidak ditemukan satupun butir yang menggunakan pernyataan berlebihan atau dianggap tidak bermanfaat. Seluruh pernyataan yang digunakan mempunyai fungsi, baik pernyataan pada rumusan pokok soal maupun pada pilihan jawaban. Sehingga peserta didik dapat memaksimalkan waktu yang disediakan untuk menjawab soal lain.

## 3. Pokok soal tidak memberi petunjuk ke arah jawaban benar

Dari 40 soal yang diuji, ditemukan 38 pokok soal yang terbebas dari mengarahkan kepada pilihan jawaban benar, sedangkan 2 soal lainnya menunjukkan ke arah jawaban benar, yaitu soal nomor 27 dan 30.

Pada soal nomor 27 terdapat sebuah bacaan yang terdiri dari satu paragraf. Untuk menegaskan maksud soal, diberikan sebuah pertanyaan. Akan tetapi jawaban secara jelas telah tercantum dalam bacaan sebelumnya pada pokok soal tersebut.

27. وَصَلَ خَالِدٌ إِلَى الطَّيِّبِ الْإِبْرَاهِيمِيِّ لِلْمُعَالَجَةِ، يُرَافِقُهُ أَخُوهُ أَحْمَدُ فَجَلَسَ خَالِدٌ فِي عُرْفَةِ الْإِنْتِظَارِ يَنْتَظِرُ دَوْرَهُ، وَ

بَعْدَ رُبْعِ سَاعَةٍ جَاءَ دَوْرُهُ، فَدَخَلَ إِلَى عُرْفَةِ الْكَشْفِ. اسْتَقْبَلَ الطَّيِّبُ خَالِدًا، فَقَالَ: "اسْتَلْقِ عَلَيَّ السَّرِيرَ!"

فَسَأَلَهُ لِيَعْرِفَ مَا يَشْكُو مِنْهُ فَقَالَ لَهُ: "مِمَّ تَشْكُو يَا أَخِي؟" فَقَالَ خَالِدٌ: "عِنْدِي أَلَمٌ فِي الرَّأْسِ وَ فِي مَعِدَّتِي

أَيْنَ جَلَسَ خَالِدٌ؟.....

A. عُرْفَةُ الْإِنْتِظَارِ

B. عُرْفَةُ الْكَشْفِ

C. عُرْفَةُ الْجُلُوسِ

D. عُرْفَةُ الْمَكْتَبِ

Pertanyaan **أَيْنَ جَلَسَ خَالِدٌ؟** (di mana kholid duduk?) mempunyai jawaban pada pilihan jawaban A, yaitu kata **عُرْفَةُ الْإِنْتِظَارِ** (ruang tunggu). Jawaban dari pertanyaan ini telah tercantum dalam paragraf yang berisi bacaan pada pokok soal yang berbunyi, فَجَلَسَ خَالِدٌ فِي عُرْفَةِ الْإِنْتِظَارِ (kemudian kholid duduk di ruang tunggu). Kalimat yang bergaris bawah tersebut secara jelas menunjukkan ke arah jawaban dari pertanyaan yang dimaksudkan.

Sedangkan pada butir soal nomor 30 yang berbunyi ..... ؟ **مَا اسْمُ هَذَا الْمَكَانِ** (apa nama tempat ini?.....), bermaksud untuk menanyakan apa nama tempat yang ditunjukkan oleh gambar tersebut dalam bahasa Arab. Akan tetapi pada gambar tersebut tampak jelas nama gambar yang dimaksud dalam bahasa Indonesia, yaitu "RUMAH SAKIT". Jika rumah sakit diterjemahkan ke dalam bahasa Arab menjadi **الْمُسْتَشْفَى** yang merujuk pada pilihan jawaban benar D. Sehingga model soal seperti ini dikategorikan ke dalam pokok soal yang memberi arahan pada pilihan jawaban yang benar melalui penegasan gambar.

#### **4. Pokok soal tidak mengandung pernyataan yang bersifat negatif ganda**

Dari 40 pokok soal, 30 berisi pernyataan positif dan 10 lainnya berisi kalimat tanya dan kalimat perintah. Pernyataan yang bersifat negatif ganda tidak dijumpai pada seluruh pokok soal. Seluruh butir soal telah sesuai dengan kaidah konstruk poin keempat dengan tidak ditemukannya pernyataan tersebut.

#### **5. Panjang rumusan pilihan jawaban relatif sama**

Pilihan jawaban dari 40 butir soal tidak ditemukan satupun pilihan jawaban yang mempunyai rumusan lebih panjang atau lebih pendek dari pilihan yang lainnya. Seluruh rumusan pilihan jawaban relatif sama panjang dan sama pendek. Sehingga pilihan jawaban sudah sesuai dengan indikator konstruk kelima, yaitu pilihan jawaban mempunyai rumusan yang hampir sama.

**6. Pilihan jawaban tidak mengandung pernyataan “semua pilihan jawaban di atas salah” atau “semua pilihan jawaban di atas benar”**

Pada indikator konstruk poin enam, tidak ditemukan kasus yang tidak sesuai. Seluruh pilihan jawaban yang tersedia mempunyai satu jawaban benar (*key*), dan yang lainnya merupakan jawaban pengecoh (*distractor*). Seluruh pilihan jawaban dari setiap butir soal telah sesuai dengan indikator konstruk ini.

**7. Pilihan jawaban yang berbentuk angka atau waktu disusun berdasarkan ukuran besar kecilnya nilai angka tersebut atau kronologisnya**

Seluruh pilihan jawaban yang terdapat pada setiap butir soal berbentuk kata dan kalimat. Tidak ditemukan satupun butir soal yang memiliki pilihan jawaban berupa angka atau waktu. Seluruh butir soal telah sesuai dengan indikator konstruk poin ini.

**8. Gambar, grafik, tabel, diagram, dan sejenisnya yang terdapat pada soal jelas, berfungsi, tidak memunculkan kebingungan, dan mempunyai tingkat keterbacaan tinggi**

Soal yang ditelaah melalui indikator ini hanyalah soal-soal yang di dalamnya terdapat gambar, grafik, tabel, diagram, dan sejenisnya. Melalui telaah kali ini, peneliti menemukan tiga butir soal yang didalamnya terdapat gambar, sedangkan 37 soal lainnya berupa kalimat saja. Tiga butir soal tersebut adalah nomor 10, 21 dan 30. Dua butir soal mempunyai gambar yang jelas dan berfungsi untuk menentukan pilihan jawaban benar serta mempunyai tingkat keterbacaan tinggi, yaitu butir nomor 10 dan 21. Sehingga kedua soal tersebut tidak menimbulkan kebingungan bagi peserta didik saat menentukan jawabannya.

Sedangkan satu butir soal, yaitu nomor 30 berisi pokok soal yang menanyakan nama dalam bahasa Arab dari gambar yang dicantumkan. Gambar tersebut menunjukkan gambar sebuah rumah sakit. Akan tetapi, pada gambar tersebut juga terdapat nama dari gambar tersebut dalam bahasa Indonesia. Hal ini menjadikan gambar tersebut kurang berfungsi. Namun gambar yang ditampilkan sudah tampak jelas, mempunyai tingkat keterbacaan tinggi serta tidak memunculkan kebingungan.

**9. Butir soal tidak bergantung pada jawaban soal sebelumnya**

Hasil telaah yang ditemukan pada indikator konstruk terakhir adalah tidak ditemukannya satupun butir soal yang mempunyai ketergantungan terhadap pilihan jawaban soal sebelumnya. Seluruh butir soal berdiri sendiri, tidak mempunyai keterkaitan dengan soal

lain, bahkan dalam hal jawaban soal sebelumnya. Sehingga tidak ditemukan satupun soal yang kurang sesuai, seluruh soal telah sesuai dengan indikator ini.

## SIMPULAN

Merujuk pada hasil pembahasan sebelumnya kita dapat menyimpulkan bahwa seluruh butir soal sesuai dengan indikator-indikator yang menunjukkan rumusan pokok soal dan pilihan jawaban hanya berisi pernyataan yang dibutuhkan, pokok soal tidak ada yang mengandung pernyataan negatif ganda, rumusan pilihan jawaban butir soal mempunyai panjang yang hampir sama, pilihan jawaban tidak mengandung pernyataan “Semua pilihan jawaban di atas salah” atau “Semua pilihan jawaban di atas benar”, dan setiap soal tidak berketergantungan dengan jawaban soal sebelumnya. Pada indikator-indikator tersebut guru telah sesuai sehingga soal sudah layak jika ditinjau dari aspek-aspek tersebut.

Di sisi lain, terdapat soal yang kurang sesuai dengan indikator yang lain, yaitu: 25 butir soal mempunyai pokok soal yang mempunyai rumusan tidak jelas dan tidak tegas, dua pokok soal yang mengarahkan pada jawaban benar, dan satu soal dengan gambar yang memunculkan kebingungan. Sehingga guru atau pembuat soal harus memperbaiki dan memperhatikan aspek-aspek tersebut agar menghasilkan alat tes yang valid serta dapat menilai secara tepat aspek yang ingin dinilai. Sebaiknya analisis validitas konstruks dilakukan oleh guru atau pembuat soal secara mandiri sebelum memberikan soal tersebut kepada peserta didiknya agar soal tepat sasaran dan mudah dipahami oleh peserta didik. Dengan demikian akan diperoleh cerminan hasil belajar yang sesuai dengan kompetensi yang dimiliki oleh peserta didik.

## DAFTAR PUSTAKA

- Al-Fathiyah, S. F. (2019). Analisis butir soal pelajaran bahasa Arab di MA Roudhotul Ulum Pagak Malang. *Tarbiyatuna: Jurnal Pendidikan Ilmiah*, 4(1), 76-100.
- Am, Z. (2018). Teknik penilaian hasil pembelajaran. *Rausyan Fikr: Jurnal Pemikiran dan Pencerahan*, 14 (2), 53-62. <http://dx.doi.org/10.31000/rf.v14i02.901>
- Antipkina, I., & Ludlow, L.H. (2020). Parental involvement as a holistic concept using Rasch/Guttman scenario scales. *Journal of Psychoeducational Assessment*, 1-20. <https://doi.org/10.1177%2F0734282920903164>

- Anwar, S. & Fakhruddin, A. (2016). Pelaksanaan standar penilaian oleh guru pendidikan agama Islam di sekolah (studi evaluatif terhadap guru PAI SMP dan SMA di Bandung). *Ta'lim: Jurnal Pendidikan Agama Islam*, 14(2), 139-155.
- Arifin, Z. (2012). *Evaluasi pembelajaran*. Jakarta: Kementerian Agama RI.
- Broadbent, J. Panadero, E., & Boud, D. (2017). Implementing summative assessment with a formative flavour: a case study in a large class. *Assessment & Evaluation in Higher Education*, 42, 1-16. <http://dx.doi.org/10.1080/02602938.2017.1343455>
- Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom, *theory into practice*, 55(2), 153-159. <http://dx.doi.org/10.1080/00405841.2016.1148989>
- Fauziah, I.R.N. (2020). Analisis kualitas tes bahasa Arab berbasis higher order thinking skill (HOTS). *Lisanuna*, 10(1), 45-54. <http://dx.doi.org/10.22373/lis.v10i1.7805>
- Furtak, E. M., Ruiz-Primo, M. A., & Bakeman, R. (2017). Exploring the utility of sequential analysis in studying informal formative assessment practices. *Educational Measurement: Issues and Practice Spring*, 36 (1), 28-38. <https://doi.org/10.1111/emip.12143>
- Hakim, R., Nugraha, A.C. (2016). Tingkat kecurangan tes *online* dibandingkan dengan tes tertulis di SMK Negeri 2 Yogyakarta, *Prodi Pendidikan Elektro: E-Journal Universitas Negeri Yogyakarta*, 6 (3), 36-41.
- Hawe, E. & Dixon, H. (2016). Assesment for learning: a catalyst for student self-regulation. *Assessment & Evaluation in Higher Education*, 42(8), 1-12. <http://dx.doi.org/10.1080/02602938.2016.1236360>
- Hikmah, M.U. (2019). Pengembangan tes sumatif bahasa Jerman semester ganjil kelas X SMA Negeri 1 Taman. *Laterne: Jurnal Mahasiswa Unesa*, 8(2), 1-21.
- Maulana, D. & Sanusi, A. (2020). Analisis butir soal bahasa Arab ujian akhir madrasah bersama daerah (UAMBD) madrasah ibtidaiyah tahun 2017-2018. *Ta'lim al-'Arabiyyah: Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, 4(1), 12-24. <https://doi.org/10.15575/jpba.v4i1.8054>
- Pratama, D. (2020). Analisis kualitas tes buatan guru melalui pendekatan *item response theory* (IRT) model rasch. *Tarbawy: Jurnal Pendidikan Islam* 7(1), 61-70. <https://doi.org/10.32923/tarbawy.v7i1.1187>

- Pusat Penilaian Pendidikan. (2019). *Panduan penilaian tertulis*. Jakarta: Pusat Penilaian Pendidikan.
- Shin, J., Bulut, O., & Griel, M.J. (2019). The effect of the most-attractive-distractor location on multiple choice item difficulty. *The Journal of Experimental Education*, 88(4), 1-17  
<https://doi.org/10.1080/00220973.2019.1629577>
- Spector, J. M., Ifenthaler, D., Sampson, D., Yang, Lan., Mukama, E., Warusavitarana, A., Dona, K.L., Eichhorn, K., Fluck, A., Huang, R., Bridges, S., Lu, J., Ren, Y., Gui, X., Deneen, C., Diego, J.S., & Gibson, D. (2016). Technology enhanced formative assessment for 21st century learning. *Educational Technology & Society*, 19(3), 58-71.
- Suharismi, A. (2011). *Dasar-dasar evaluasi pendidikan*. Jakarta: Bumi Aksara.
- Syihabuddin. (2018). *Tes dan evaluasi pengajaran bahasa*. Bandung: UPI Press.
- Tempelaar D., Rienties, B., Mittelmeier, J., & Nguyen, Q. (2017). Student profiling in a dispositional learning analytics application using formative assessment. *Computers in Human Behavior*, 78, 1-13. <https://doi.org/10.1016/j.chb.2017.08.010>