

BAB II

KAJIAN PUSTAKA

2.1. ALAT EVALUASI

Dalam proses evaluasi diperlukan adanya sebuah alat. Alat yang dimaksudkan adalah alat yang berfungsi untuk memudahkan pada saat proses evaluasi. Berikut ini akan dijelaskan secara rinci dan sistematis terkait alat evaluasi.

2.1.1. Pengertian Alat Evaluasi

Kata “alat” biasa disebut juga dengan istilah “instrumen”. Dalam pengertian umum menurut Kamus Besar Bahasa Indonesia (KBBI) alat adalah benda yang dipakai untuk mengerjakan sesuatu. Sedangkan menurut . (Arikunto, 2016) alat adalah sesuatu yang dapat digunakan untuk mempermudah seseorang dalam melaksanakan tugas atau mencapai tujuan secara lebih efektif dan efisien. Dalam kegiatan evaluasi, fungsi alat adalah untuk mengetahui dan memperoleh hasil belajar.

Menurut Permendikbud Nomor 104 Tahun 2014 penilaian hasil belajar (evaluasi) adalah proses pengumpulan informasi/bukti tentang capaian pembelajaran peserta didik dalam kompetensi sikap spiritual dan sikap sosial, kompetensi pengetahuan, dan kompetensi keterampilan yang dilakukan secara terencana dan sistematis, selama dan setelah proses pembelajaran. Alat evaluasi dapat berupa lisan atau tulisan dan sangat diperlukan untuk bisa mengetahui kemampuan berpikir peserta didik atas kegiatan pembelajaran yang telah dilakukan. Hal ini di tandai dengan kemampuan peserta didik mengerjakan soal evaluasi tersebut.

Menurut (Hamzah, 2014) instrumen evaluasi pembelajaran matematika adalah alat ukur yang dipakai dalam pembelajaran, untuk menilai dan mengevaluasi sampai sejauh mana proses pembelajaran matematika mencapai sasarannya. Banyak alat atau instrument yang dapat digunakan dalam kegiatan evaluasi hasil pembelajaran, salah satunya adalah tes (Arifin, 2013). Alat ukur yang dimaksudkan tersebut tidak dapat dibandingkan antara satu sekolah

dengan sekolah lainnya. Dan alat evaluasi dikatakan baik apabila mampu mengevaluasi sesuatu dengan hasil seperti keadaan yang dievaluasi.

Jadi, dari berbagai pendapat diatas dapat disimpulkan instrument evaluasi atau alat evaluasi adalah sebuah alat ukur yang dalam berbagai bentuk, baik tulisan, atau lisan untuk mengukur hasil belajar.

2.1.2. Teknik-Teknik Evaluasi

Menurut (Arikunto, 2016) teknik evaluasi dibagi menjadi 2 yakni teknik tes dan teknik non-tes.

2.1.2.1. Teknik Tes

Tes merupakan suatu alat pengumpul informasi, tetapi jika dibandingkan dengan alat-alat yang lain, tes bersifat lebih resmi karena penuh dengan batasan- batasan. Tes mempunyai fungsi ganda, yaitu untuk mengukur peserta didik dan untuk mengukur keberhasilan program pengajaran (Arikunto, 2016). Sedangkan menurut (Hamzah, 2014) tes adalah alat atau prosedur yang digunakan untuk mengetahui atau mengukur sesuatu dalam suasana dengan cara dan aturan-aturan yang sudah ditentukan. Selain itu menurut (Arifin, 2013) tes banyak digunakan untuk mengukur prestasi belajar peserta didik dalam bidang kognitif, seperti pengetahuan, pemahaman, aplikasi, analisis, dan yang lainnya. Sehingga sering kali dalam pembelajaran, tes dikaitkan dengan penilaian kognitif peserta didik.

Menurut (Arikunto, 2016) Tes juga sudah dikembangkan menjadi beberapa bentuk, yaitu :

1. Tes subjektif

Menurut (Arikunto, 2016) tes subjektif merupakan tes yang pada umumnya berbentuk esai (uraian). Disebut bentuk uraian, karena menuntut peserta didik untuk menguraikan, mengorganisasikan dan menyatakan jawaban dengan kata-katanya sendiri dalam bentuk, teknik, dan gaya yang berbeda satu dengan lainnya. Tes subjektif ini memiliki kelebihan, yaitu mudah disusun dan disiapkan, dan dapat diketahui sejauh mana peserta didik mendalami suatu materi yang diteskan. Selain memiliki kelebihan, tes uraian juga memiliki kekurangan, yaitu kadar vasilitas dan reliabilitas rendah, cara memeriksanya banyak dipengaruhi oleh unsur-unsur subjektif, membutuhkan waktu yang lama dalam mengoreksi.

2. Tes objektif

Menurut (Arikunto, 2016) tes objektif merupakan tes yang dalam pemeriksaannya dapat dilakukan secara objektif. Disebut tes objektif karena menurut (Arikunto, 2016) bentuk penilaiannya objektif, siapapun yang mengoreksi jawaban tes objektif hasilnya akan sama karena kunci jawabannya sudah jelas dan pasti. Dalam tes objektif ada beberapa macam, diantaranya adalah:

- a) True or False : Menurut (Arikunto, 2016) merupakan tes objektif yang soal – soalnya berupa pernyataan yang mengandung dua kemungkinan jawaban, yaitu benar atau salah. Peserta didik harus memilih salah satu dari kemungkinan jawaban atas pernyataan tersebut.
- b) Multiple choice : Menurut (Arikunto, 2016) merupakan tes objektif yang terdiri atas suatu keterangan atau pemberitahuan tentang suatu pengertian yang belum lengkap. Dan untuk melengkapinya harus memilih satu jawaban dari beberapa kemungkinan jawaban yang telah disediakan.
- c) Matching : Menurut (Arikunto, 2016) merupakan tes objektif yang terdiri atas satu seri pertanyaan dan satu seri jawaban. Tugas peserta didik mencari, memasangkan atau mencocokkan jawaban-jawaban sehingga sesuai dengan pertanyaannya.
- d) Fill in the blank : Menurut (Arikunto, 2016) merupakan tes objektif yang terdiri atas kalimat-kalimat yang ada bagian-bagiannya yang dihilangkan. Bagian yang hilang tersebutlah yang harus diisi oleh peserta didik.

2.1.2.2. Teknik Non-tes

Teknik nontes merupakan salah satu teknik yang digunakan dalam evaluasi hasil belajar peserta didik namun tanpa pengujian. Teknik nontes ini biasanya digunakan oleh pendidik dalam penilaian ranah afektif (sikap) dan ranah psikomotorik (keterampilan). Karena teknik nontes sulit digunakan untuk penilaian kognitif. Teknik penilaian non tes dapat dikelompokkan lima menurut (Arikunto, 2016) sebagai berikut :

- a) Skala bertingkat

Skala menggambarkan suatu nilai yang berbentuk angka terhadap sesuatu hasil pertimbangan. Seperti Oppenheim mengatakan: *Rating gives a*

numerical value to some kind of judgement, maka suatu skala selalu disajikan dalam bentuk angka (Arikunto, 2016).

b) Kuesioner

Kuesioner juga sering dikenal dengan sebagai angket. Kuesioner adalah sebuah daftar pertanyaan yang harus diisi oleh orang yang akan diukur (Responden) (Arikunto, 2016). Berdasarkan siapa yang mengisi kuesioner maka dikenal kuesioner langsung dan kuesioner tak langsung. Kuesioner langsung adalah seperangkat pertanyaan yang dikirim langsung kepada orang yang diminta pendapat. Contoh dari kuesioner langsung adalah guru matematika yang menyebarkan kuesioner langsung kepada peserta didik dikelas untuk mengetahui daya serap peserta didik atas penjelasan yang diberikan. Sedangkan dikatakan angket tak langsung karena ketika ingin mendapatkan data dari responden yang diberi kuesioner adalah orang lain, misalnya ingin mengetahui kebiasaan anak belajar dirumah, disebarkan kuesioner kepada orang tua.

Keuntungan dari angket yaitu: (1) responden dapat menjawab dengan bebas tanpa dipengaruhi oleh hubungan dengan peneliti atau penilai, dan waktu relatif lama, sehingga objektivitas terjamin (2) informasi atau data terkumpul lebih mudah karena itemnya homogen (3) dapat digunakan untuk mengumpulkan data dari jumlah responden yang besar yang dijadikan sampel. Sedangkan kelemahannya adalah (1) ada kemungkinan angket diisi oleh orang lain (2) hanya diperuntukkan bagi yang dapat melihat saja (3) responden hanya menjawab yang ada (Arifin, 2013).

c) Daftar cocok

Yang dimaksud dengan daftar cocok adalah deretan pernyataan (yang biasanya singkat-singkat), dimana responden yang dievaluasi tinggal membubuhkan tanda cocok ditempat yang sudah disediakan (Arikunto, 2016).

d) Wawancara

Wawancara adalah suatu metode atau cara yang digunakan untuk mendapatkan jawaban dari responden dengan cara tanya-jawab sepihak (Arikunto, 2016). Dikatakan sepihak karena dalam wawancara ini responden tidak diberi kesempatan sama sekali untuk mengajukan pertanyaan. Secara umum, dalam (Sudijono, 2015) yang dimaksud wawancara adalah cara menghimpun data-data

keterangan yang dilaksanakan dengan cara melakukan tanya jawab lisan secara tatap muka dan dengan arah tujuan yang telah ditentukan.

Menurut (Arifin, 2013) tujuan wawancara adalah sebagai berikut: (1) untuk memperoleh informasi secara langsung guna menjelaskan suatu hal atau situasi dan kondisi tertentu. (2) untuk melengkapi suatu penyelidikan ilmiah. (3) untuk memperoleh data agar dapat memengaruhi situasi atau orang tertentu. Wawancara yang dilakukan pendidik misalnya menanyakan kepada peserta didik tentang bagaimana pendapatnya atas penyampaian materi guru tersebut, atau tentang kesulitan-kesulitan yang dihadapi oleh peserta didik dalam memahami materi.

e) Pengamatan

Pengamatan atau observasi adalah teknik penilaian yang dilakukan oleh pendidik dengan menggunakan indra secara langsung. Menurut (Arikunto, 2016) ada 3 macam observasi: (1) observasi partisipan, yaitu observasi yang dilakukan oleh pengamat, dalam hal ini pengamat memasuki dan mengikuti kegiatan kelompok yang sedang diamati. (2) observasi sistematis, yaitu observasi dimana faktor-faktor yang diamati sudah didaftar secara sistematis dan sudah diatur menurut kategorinya. (3) observasi eksperimental terjadi jika pengamat tidak berpartisipasi dalam kelompok.

Contoh pengamatan yang bisa dilakukan dalam pelajaran matematika misalnya sikap ketelitian, ketekunan, dan kecepatan dalam penyelesaian soal.

2.2. KUALITAS BUTIR SOAL

Tes evaluasi yang digunakan dalam sebuah alat evaluasi perlu dilihat kualitasnya. Menurut (Arikunto, 2016) mengemukakan bahwa analisis kualitas tes merupakan kegiatan untuk mengkaji soal pada setiap item atau butirnya guna mengetahui kualitas dari setiap butir soal tersebut. Untuk mengetahui kualitas butir soal dapat dilakukan dengan dua cara yaitu secara kualitatif dan kuantitatif. Dalam penelitian ini akan dilakukan analisis kualitas butir soal secara kuantitatif yang berkaitan dengan ciri – ciri statistiknya. Analisis secara kuantitatif ini juga disebut dengan istilah analisis empirik. Analisis ini dilakukan dengan mengukur beberapa kriteria kualitas soal, diantaranya adalah validitas, reliabilitas, tingkat kesukaran, daya beda dan efektivitas pengecoh.

2.2.1. Validitas

Validitas berasal dari kata *validity* yang artinya sejauh mana ketepatan dan kecermatan suatu alat ukur dalam melakukan fungsi ukurnya (Djaali., 2008) Menurut (Arikunto, 2016) Validitas adalah suatu ukuran yang menunjukkan tingkat-tingkat kevalidan atau keshahihan suatu instrumen. Alat evaluasi yang memiliki tingkat validitas tinggi maka alat evaluasi tersebut bisa dikatakan valid, begitupun sebaliknya, alat evaluasi tidak bisa dikatakan valid atau kurang valid jika memiliki tingkat validitasnya rendah.

2.2.2. Reliabilitas

Menurut (Arikunto, 2016) pengertian dari reliabilitas tes berhubungan dengan masalah ketepatan hasil. Suatu hasil pengukuran hanya dapat dipercaya apabila dalam beberapa kali pelaksanaan pengukuran terhadap kelompok yang sama diperoleh hasil pengukuran yang relatif sama selama aspek yang diukur dalam diri subjek memang belum berubah (Hamzah, 2014).

2.2.3. Tingkat Kesukaran

Menurut (Hamzah, 2014) Salah satu karakteristik butir soal yang dapat menunjukkan kualitas butir soal tersebut termasuk dalam kategori sukar, sedang, atau mudah. Butir soal dikatakan sukar jika banyak peserta didik yang menjawab soal tersebut salah. Sebaliknya, jika banyak peserta didik yang menjawab suatu butir soal benar maka dapat dikatakan butir soal tersebut termasuk dalam kategori mudah.

Menurut (Arikunto S. , 2006) tingkat kesukaran adalah presentase jumlah peserta didik yang menjawab soal dengan benar atau salah. Maka Jika suatu soal memiliki tingkat kesukaran seimbang (Proporsional), maka dapat dikatakan bahwa soal tersebut baik. sehingga sebaiknya soal tes termasuk dalam kategori sedang, tidak terlalu mudah dan tidak terlalu sulit.

Jadi dari beberapa definisi tersebut, dapat disimpulkan bahwa tingkat kesukaran merupakan salah satu karakteristik butir soal yang harus diperhatikan dalam menganalisis butir soal untuk mengetahui seberapa besar tingkat kesukaran dalam setiap soal.

2.2.4. Daya Pembeda

Menurut (Arikunto S. , 2016) yang berpendapat bahwa daya pembeda adalah kemampuan suatu soal untuk membedakan antara peserta didik yang pandai (berkemampuan tinggi) dengan peserta didik yang bodoh (berkemampuan rendah). Semakin tinggi daya beda suatu soal maka semakin baik soal tersebut untuk membedakan antara peserta didik yang berkemampuan tinggi dengan peserta didik yang berkemampuan rendah. Menurut (Hamzah, 2014) dalam penyusunan butir soal tes, sebaiknya ada sifat yang menunjukkan kualitasnya sehingga :

1. Tidak dapat dijawab benar baik oleh peserta didik kelompok atas maupun peserta didik kelompok bawah.
2. Dapat dijawab benar oleh peserta didik kelompok atas tapi tidak dapat dijawab oleh peserta didik kelompok bawah.
3. Dapat dijawab benar oleh peserta didik kelompok atas maupun kelompok bawah.

2.2.5. Efektifitas Pengecoh

Menurut (Arifin, 2013) pengecoh akan dianggap baik jika jumlah peserta didik yang memilih pengecoh itu sama atau mendekati jumlah ideal. Butir soal yang baik adalah jika pengecohnya akan dipilih secara merata oleh peserta didik yang menjawab salah. Sebaliknya, butir soal yang kurang baik, jika pengecohnya akan dipilih secara tidak merata oleh peserta didik.

2.3. TWO TIER MULTIPLE CHOICE

Dalam proses pembelajaran, terdapat berbagai macam tes yang digunakan. Bentuk tes yang sering digunakan adalah essay (uraian) dan *multiple choice* (pilihan ganda). Masing-masing dari bentuk tes yang sering digunakan ini memiliki kelebihan dan kekurangan.

Bentuk tes essay, memberikan kebebasan kepada setiap penempuh tes untuk mengekspresikan daya nalarnya, sehingga jawaban yang diberikan oleh setiap penempuh tes akan menunjukkan kemampuan berpikir secara kompleks (Smith, 2020). Selain itu bentuk soal essay juga meminimalisir peserta didik yang asal menjawab. Karena kebanyakan peserta didik yang diberikan soal essay apabila tidak bisa menjawab maka akan mengosongi jawabanya. Namun demikian ada

kelemahan bentuk tes essay. Bentuk tes essay dalam memberikan skor membutuhkan waktu yang lama dan relatif lebih sulit. Di samping itu, penskoran bentuk tes essay bersifat subjektif dan harus dilakukan oleh ahli atau yang berwenang sehingga tidak dapat dilakukan komputersasi dalam penskorannya (Smith, 2020).

Berbeda dengan bentuk tes essay, bentuk tes *multiple choice* lebih praktis dalam penskorannya. Pada bentuk tes *multiple choice* siapa pun yang memeriksa akan memberikan skor yang sama, sehingga kesalahan dalam penskoran dapat menjadi kecil. Hal ini juga ditegaskan oleh (Sudijono, 2015) bahwa bentuk tes *multiple choice* memberikan hasil yang objektif yang tinggi dan dapat dipercaya, dalam pengoreksian jawabannya tidak sulit, dan lebih representatif dalam hal mencakup dan mewakili materi yang telah diajarkan. Menurut sebagian banyak orang, tes bentuk seperti ini lebih efisien penggunaannya dalam proses evaluasi pembelajaran karena bisa mencakup materi pembelajaran yang lebih luas. Hal ini sejalan dengan sebuah pendapat yang dikemukakan oleh (Arikunto S. , 2016) bahwa tes bentuk pilihan ganda merupakan bentuk tes objektif yang paling banyak digunakan karena banyak sekali materi yang dapat dicakup.

Namun demikian bentuk tes *multiple choice* mempunyai peluang menjawab benar dengan menebak cukup tinggi. Selain itu, bentuk tes objektif *multiple choice* atau pilihan ganda hanya terbatas pada mengetahui pilihan jawaban peserta didik saja. Maka sebaiknya pendidik menyusun tes pilihan ganda yang meminta penjelasan atau alasan dari peserta didiknya dalam menjawab. Tes pilihan ganda yang dilengkapi dengan alasan itu termasuk dalam jenis tes diagnostik. Hal ini relevan dengan pendapat menurut (Susanti, 2014) bahwa Tes diagnostik dapat berupa, tes *multiple choice* dengan *reasoning* terbuka, tes *multiple choice* dengan alasan yang sudah ditentukan dan tes esai tertulis.

Hasil perbaikan atau modifikasi bentuk tes objektif pilihan ganda yang dilengkapi dengan alasan adalah *two tier multiple choice*. *Two tier multiple choice* yaitu tes berupa soal pilihan ganda dengan dua tingkat, pertama kali dikembangkan oleh Dafid F. Treagus. Pada tingkat pertama soal tentang pertanyaan perihal rancangan yang akan diuji cobakan sedangkan pada tingkat kedua tentang alasan setiap jawaban yang ada pada pertanyaan pada tingkat

pertama. Menurutnya hal ini dianggap lebih efektif untuk digunakan pada evaluasi pembelajaran. Karena Bentuk soal objektif berupa *two tier multiple choice* juga dapat mengurangi peserta didik dalam menjawab pertanyaan dengan cara menebak, sebab bentuk soal yang seperti ini mengharuskan peserta didik untuk memberikan alasan yang sesuai dengan jawaban yang telah dipilihnya.

TTMC atau singkatan dari *two tier multiple choice* menurut (Junari, 2017) merupakan hasil perbaikan dari tes pilihan ganda yang secara luas telah banyak digunakan dalam kegiatan evaluasi pembelajaran. Kelemahan tes pilihan ganda yang kurang mampu menggambarkan kemampuan peserta didik menjadi cikal bakal dikembangkan TTMC. Penggunaan TTMC lebih mampu memberikan informasi yang meyakinkan tentang pemahaman konsep peserta didik dibanding tes pilihan ganda biasa.

Sedangkan menurut (Rositasari, 2014) *two tier multiple choice* memiliki kelebihan:

- a. Menurunkan kemampuan menebak. Pada tes pilihan ganda biasa, apabila terdapat lima pilihan jawaban maka setiap pilihan jawaban memiliki 20% kemungkinan jawaban benar. Sedangkan pada tes pilihan ganda *Two-tier*, apabila terdapat lima pilihan memiliki 4% kemungkinan untuk mendapatkan jawaban benar.
- b. Memungkinkan menggabung beberapa aspek dalam suatu fenomena. Pada *tier* pertama, siswa diminta untuk menjawab pertanyaan. Sedangkan pada *tier* kedua, peserta didik diminta untuk memberikan alasan dari jawaban tersebut. Oleh karena itu, dapat dikatakan bahwa *tier* pertama merupakan *menological domain* sedangkan *tier* kedua merupakan *conceptual domain*.
- c. *Two tier multiple choice* memiliki kelebihan dari pada pilihan ganda biasa pada umumnya. Tes diagnosis *two tier* lebih mudah dikelola dan dihitung dibandingkan dengan metode lainnya, dengan demikian sangat berguna digunakan dalam kelas.

Selain kelebihan-kelebihan yang sudah dijelaskan di atas, soal berbentuk *two tier multiple choice* dapat dikatakan lebih baik dari soal bentuk essay jika memiliki pengocoh yang baik dan juga dengan daya pembeda yang baik. Hal ini juga sependapat dengan (Wulandari, 2015) yang menyatakan bahwa kualitas suatu

soal berbentuk *two tier multiple choice* dapat dikatakan baik dengan memperhatikan beberapa indikator kualitas soal yang digunakan dalam tes, terutama pada daya pembeda dan pengecoh.

Dalam menyusun soal bentuk *two tier multiple choice* yang telah dikembangkan oleh Dafid F. Treagus ada prosedur–prosedur yang harus dilakukan, seperti yang dijabarkan oleh (Rositasari, 2014) prosedur–prosedur dalam menyusun soal bentuk *two tier multiple choice* antara lain :

1. Tahap pendefinisian (*Defining the content*)

- a. Mengidentifikasi pernyataan-pernyataan mengenai pengetahuan konsep (*Identifying propositional knowledge statements*)
- b. Mengembangkan peta konsep (*Development a concept map*)
- c. Menghubungkan pernyataan-pernyataan mengenai pengetahuan konsep dengan peta konsep (*Relating propotional knowledge to the concept map*)

Pernyataan-pernyataan mengenai pengetahuan konsep dihubungkan secara langsung dengan pea konsep untuk memastikan bahwa pernyataan-pernyataan mengenai pengetahuan konsep yang telah dianalisis sejalan dengan peta konsep.

d. Memvalidasi isi (*Validating the content*)

Tahapan ini ialah memvalidasi pernyataan mengenai pengetahuan konsep dan peta konsep yang telah dianalisis dan dibuat. Validasi dilakukan oleh pendidik sains, dosen, dan ilmuwan atau spesialis sains dengan pengetahuan menyeluruh tentang materi.

2. Memperoleh informasi mengenai kesalahan peserta didik (*Obtaining information about students' misconceptions*)

- a. Memeriksa kepustakaan atau penelitian yang bersangkutan (*Examining relate literature*)

Manfaat dari tahapan ini adalah sebagai awalan untuk membangun informasi dasar dalam mengembangkan pertanyaan pilihan ganda untuk mengidentifikasi kesalahan pada topik tersebut.

- b. Mengadakan tanya-jawab tidak terstruktur (*Conducting unstructured student interviews*)

Untuk memperoleh lebih luas lagi pemahaman peserta didik mengenai topik yang sedang diinvestigasi, tanya jawab tidak terstruktur dilakukan kepada peserta didik yang sudah pernah mempelajari mengenai topik tersebut.

- c. Mengembangkan soal pilihan ganda dengan alasan terbuka
(*Developing multiple choice content items with free response*)

Setiap soal yang dikembangkan berdasarkan topik yang telah dipelajari, terdiri dari item soal dari konsep terkait, setiap soal pilihan ganda diikuti dengan alasan. Setelah soal-soal dibuat, kemudian diteskan kepada suatu kelas.

3. Pengembangan tes diagnostik (*Developing a diagnostic test*)

- a. Pengembangan tes diagnostik *two tier* (*Developing two tier diagnostic test*)

Tes diagnostik *two tier* terdiri dari *tier* pertama yaitu pilihan ganda jawaban, sedangkan *tier* kedua merupakan pilihan ganda alasan yang merupakan kemungkinan alasan dari jawaban *tier* pertama.

- b. Membuat kisi-kisi tes (*Designing a specification grid*)

Tahapan ini bertujuan untuk memastikan bahwa pengujian tes diagnostic dapat mencakup konsep-konsep pengetahuan dan konsep-konsep dalam peta konsep yang mendasari topik.

- c. Melanjutkan penyempurnaan (*Continuing refinements*)

Perbaikan pilihan ganda *two tier* untuk membuat pilihan ganda *two-tier* dapat menguji keseluruhan dengan baik untuk digunakan dalam tes.

Jadi, berdasarkan uraian diatas dapat disimpulkan bahwa soal bentuk TTMC atau *two tier multiple choice* merupakan tes objektif hasil perbaikan dari tes pilihan ganda. Tes ini terdiri dari 2 tingkatan. Tingkatan pertama berisi pertanyaan yang mengandung berbagai pilihan jawaban, sedangkan pada tingkatan kedua berisi alasan untuk setiap jawaban pada tingkat pertama. Selain itu, *two tier multiple choice* juga merupakan salah satu bentuk tes diagnostik yang dapat digunakan untuk mengetahui tingkat kemampuan berpikir tingkat tinggi peserta didik.

Kriterian pemberian skor pada tes diagnostik *two tier multiple choice* (pilihan ganda tingkat dua) adalah sebagai berikut :

- a. Jika peserta didik memilih jawaban dan alasan benar maka skor = 1
- b. Jika peserta didik memilih jawaban benar dan alasan salah maka skor = 0
- c. Jika peserta didik memilih jawaban salah dan alasan benar maka skor = 0
- d. Jika peserta didik memilih jawaban dan alasan salah maka skor = 0

Tabel 2.1 Keterkaitan Kriteria Dengan Jawaban

N0	Kriteria jawaban peserta didik	Kriteria
1.	Jawaban benar – alasan benar	Baik
2.	Jawaban benar – alasan salah	Cukup baik
3.	Jawaban salah – alasan benar	Cukup baik
4.	Jawaban salah – alasan salah	Kurang baik

2.4. Higher Order Thinking Skills (HOTS)

Higher Order Thinking Skill (HOTS) atau kemampuan berpikir tingkat tinggi dijelaskan oleh Rosnawati dalam (Laily, 2015) menjelaskan kemampuan berpikir tingkat tinggi dapat terjadi ketika seseorang mengaitkan informasi yang baru diterima dengan informasi yang sudah tersimpan di dalam ingatannya, kemudian menghubung-hubungkannya dan/atau menata ulang serta mengembangkan informasi tersebut sehingga tercapai suatu tujuan ataupun suatu penyelesaian dari suatu keadaan yang sulit dipecahkan.

HOTS seringkali dijabarkan sebagai keterampilan berpikir level tinggi pada berbagai kerangka keterampilan berpikir. Dalam hal ini, istilah HOTS biasanya dikontraskan dengan LOTS (*lower order thinking skill*). Dalam (Krathwohl, 2010) mengelompokkan proses kognitif analisis (*analysis*), sintesis (*synthesis*), dan evaluasi (*evaluation*) dalam taksonomi Bloom sebagai HOTS, sedangkan pengetahuan (*knowledge*), pemahaman (*comprehension*) dan penerapan (*application*) sebagai LOTS.

Taksonomi Bloom pada ranah kognitif merupakan dasar bagi keterampilan berpikir tingkat tinggi atau dikenal dengan istilah *Higher Order Thinking Skills* (HOTS). Tingkatan taksonomi Bloom pada awalnya yakni: (1) pengetahuan

(*knowledge*); (2) pemahaman (*comprehension*); (3) penerapan (*application*); (4) analisis (*analysis*); (5) sintesis (*synthesis*); dan (6) evaluasi (*evaluation*).

Revisi dilakukan terhadap taksonomi Bloom, yakni perubahan dari kata benda menjadi kata kerja. Perubahan ini dibuat agar sesuai dengan tujuan-tujuan pendidikan yang mengindikasikan bahwa peserta didik akan dapat melakukan sesuatu (kata kerja) dengan sesuatu (kata benda). Revisi dilakukan oleh Krathwohl dan Anderson, taksonomi Bloom menjadi: (1) mengingat (*remember*); (2) memahami (*understand*); (3) mengaplikasikan (*apply*); (4) menganalisis (*analysis*); (5) mengevaluasi (*evaluate*); dan (6) mencipta/mengkreasi (*create*). Kemampuan yang melibatkan menganalisis, mengevaluasi dan mencipta/mengkreasi inilah yang dinamakan ketrampilan berpikir tingkat tinggi atau dikenal dengan istilah *Higher Order Thinking Skills* (HOTS).

Menurut (Krathwohl, 2010) menyatakan bahwa indikator untuk mengukur kemampuan berpikir tingkat tinggi meliputi:

1. Menganalisis

(a) Menganalisis informasi yang masuk dan membagi-bagi atau menstrukturkan informasi ke dalam bagian yang lebih kecil untuk mengenali pola atau hubungannya. (b) Mampu mengenali serta membedakan faktor penyebab dan akibat dari sebuah skenario yang rumit. (c) Mengidentifikasi/merumuskan pertanyaan.

2. Mengevaluasi

(a) Memberikan penilaian terhadap solusi, gagasan dan metodologi dengan menggunakan kriteria yang cocok atau standar yang ada untuk memastikan nilai efektivitas atau manfaatnya. (b) Membuat hipotesis, mengkritik dan melakukan pengujian. (c) Menerima atau menolak suatu pernyataan berdasarkan kriteria yang telah ditetapkan.

3. Mencipta/mengkreasi

(a) Membuat generalisasi suatu ide atau cara pandang terhadap sesuatu. (b) Merancang suatu cara untuk menyelesaikan masalah. (c) Mengorganisasikan unsur-unsur atau bagian-bagian menjadi struktur baru yang belum pernah ada sebelumnya.

Adapun langkah-langkah penyusunan soal HOTS menurut (Kemendikbud, 2019) adalah sebagai berikut:

1. Menganalisis KD

Analisis KD diawali dengan menentukan KD yang terdapat pada Permendikbud no. 37 tahun 2018. Selanjutnya, KD yang sudah ditentukan dianalisis berdasarkan tingkat kognitifnya. Tidak semua KD yang terdapat pada Permendikbud no. 37 tahun 2018 berada dalam tingkat kognitif yang sama. KD yang berada pada tingkat kognitif C4 (menganalisis), C5 (mengevaluasi), dan C6 (mengkreasikan) dapat disusun soal HOTS. KD yang berada pada tingkat kognitif C1 (mengingat), C2 (memahami), dan C3 (menerapkan) tidak dapat langsung disusun soal HOTS. KD tersebut dapat disusun soal HOTS, bila sebelumnya dirumuskan terlebih dahulu IPK pengayaan dengan tingkat kognitif C4, C5, dan C6.

2. Menyusun kisi-kisi soal

Kisi-kisi penyusunan soal digunakan pendidik untuk menyusun soal HOTS diantaranya meliputi: (a) Memilih KD yang dapat dibuat soal HOTS. (b) Menentukan lingkup materi dan materi yang terkait dengan KD yang akan diuji. (c) Merumuskan indikator soal. (d) Menentukan nomor soal. (e) Menentukan level kognitif (L1 untuk tingkat kognitif C1 dan C2, L2 untuk tingkat C3, dan L3 untuk tingkat kognitif C4, C5, dan C6). (f) Menentukan bentuk soal yang akan digunakan.

3. Memilih stimulus yang tepat dan kontekstual

Stimulus yang digunakan harus tepat, artinya mendorong peserta didik untuk mencermati soal. Stimulus yang tepat umumnya baru dan belum pernah dibaca oleh peserta didik. Stimulus kontekstual dimaksudkan stimulus yang sesuai dengan kenyataan dalam kehidupan sehari-hari, menarik, mendorong peserta didik untuk membaca. Dalam konteks Ujian Sekolah, guru dapat memilih stimulus dari lingkungan sekolah atau daerah setempat. Stimulus juga bisa berupa gambar atau video.

4. Menulis butir pertanyaan sesuai dengan kisi-kisi soal

Butir-butir pertanyaan ditulis sesuai dengan kaidah penulisan butir soal HOTS. Kaidah penulisan butir soal HOTS, agak berbeda dengan kaidah penulisan butir soal pada umumnya. Perbedaannya terletak pada aspek materi, sedangkan

pada aspek konstruksi dan bahasa relatif sama. Setiap butir soal ditulis pada kartu soal.

5. Membuat pedoman penskoran (rubrik) atau kunci jawaban

Setiap butir soal HOTS yang ditulis hendaknya dilengkapi dengan pedoman penskoran atau kunci jawaban. Pedoman penskoran dibuat untuk bentuk soal uraian. Sedangkan kunci jawaban dibuat untuk bentuk soal pilihan ganda, pilihan ganda kompleks (benar/salah, ya/tidak), dan isian singkat.

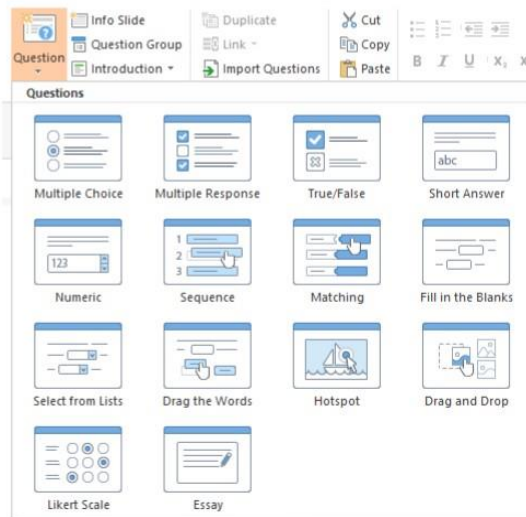
2.5. ISPRING SUITE 9

Ispring suite 9 merupakan *software* yang dapat diintegrasikan dengan *microsoft power point*, sehingga tidak begitu sulit dalam menggunakan *software* ini dan dalam penyimpanan hasil dari *software* ini dapat berupa file berbentuk *flash*, video atau lainnya. Menurut (Cahyati, 2018) *ispring suite 9* merupakan salah satu tool yang mengubah file presentasi menjadi bentuk *flash* dan bentuk SCORM / AICC, yaitu bentuk yang biasa digunakan dalam pembelajaran dengan *e-learning LMS (Learning Management System)*. Perangkat lunak *ispring* tersedia dalam versi *free* (gratis) dan berbayar.

Software ispring yang paling terbaru adalah *Ispring suite 9* yang baru saja dikembangkan dan dirilis pada April 2018. *software* ini dapat membantu dan memudahkan pendidik dalam proses pengajarannya agar lebih menarik dan efektif dengan fasilitas, fitur, dan kelebihan-kelebihan yang ada. Berikut ini adalah kelebihan-kelebihan dari *software ispring suite 9* :

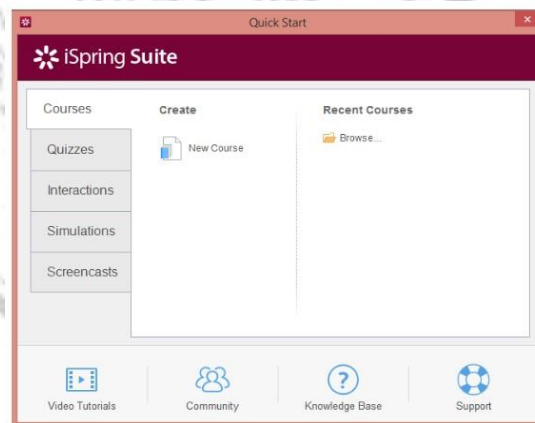
- a. *Ispring suite 9* bekerja sebagai *add-ins PowerPoint*, untuk menjadikan file *PowerPoint* lebih menarik dan interaktif berbasis *Flash* dan dapat dibuka di hampir setiap komputer atau *platform*.
- b. Dikembangkan untuk mendukung *e-learning*. *Ispring suite 9* dapat menyisipkan berbagai bentuk media, sehingga media pembelajaran yang dihasilkan akan lebih menarik, diantaranya adalah dapat merekam dan sinkronisasi video presenter, menambahkan *Flash* dan video *YouTube*, mengimpor atau merekam audio, menambahkan informasi pembuat presentasi dan logo perusahaan, serta membuat navigasi dan desain yang unik.
- c. Mudah didistribusikan dalam format *Flash*, yang dapat digunakan dimanapun dan dioptimalkan untuk web.

- d. Membuat kuis dengan berbagai jenis bentuk pertanyaan/soal yaitu : *Multiple Choice* (Pilihan Ganda), *Multiple response* (Respon ganda), *True Or False* (Benar atau Salah), *Short Answer* (Jawaban Pendek), *Numeric* (Angka), *Sequences* (Mengurutkan), *Matching* (Menjodohkan), *Fill In the Blank* (Mengisi jawaban yang kosong), dan yang lainnya. Berikut ini gambar atau bentuk soal yang dapat digunakan di *ispring suite 9*.



Gambar 2.1 Macam-Macam Bentuk Soal

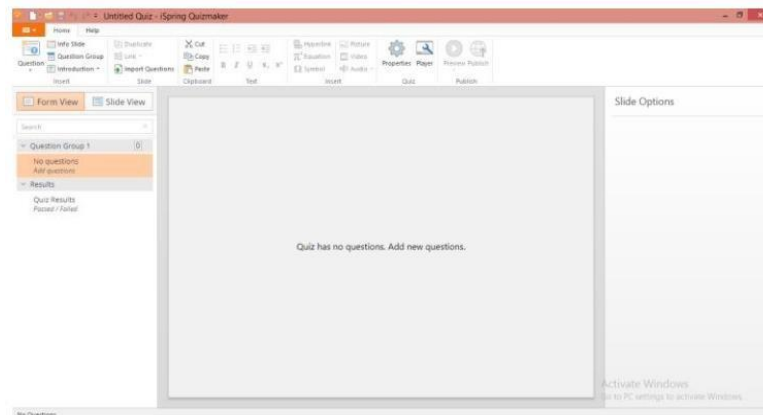
Ispring suite 9 ini dapat digunakan untuk beberapa hal. Dapat diketahui pada halaman utama dari *ispring suite 9*.



Gambar 2.2 Halaman Utama *Ispring Suite 9*

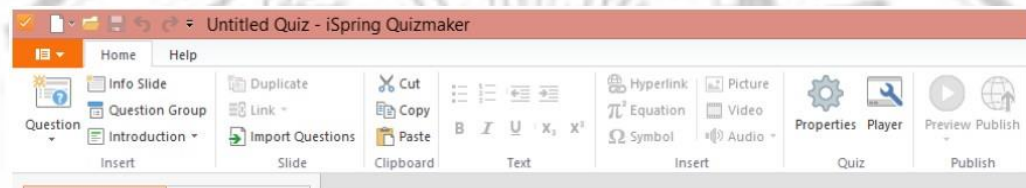
Ada 5 media yang dapat dibuat dalam *ispring suite 9*. Namun, dalam membuat sebuah alat evaluasi media yang digunakan adalah media *quizzes*. Media *quizzes* merupakan media yang dapat digunakan di *ispring suite 9* untuk membuat kuis. Media kuis ini dapat digunakan oleh pendidik

saat melakukan evaluasi pembelajaran. Berikut ini tampilan utama pada media kuis di *ispring suite 9*.



Gambar 2.3 Halaman Utama Kuis *Ispring Suite 9*

Pada media kuis *ispring suite 9* ini ada beberapa menu-menu utama, diantaranya adalah :



Gambar 2.4 Menu Utama Media Kuis *Ispring Suite 9*

1. *Insert 1*: digunakan untuk menyisipkan dan mengatur berbagai bentuk soal, dan slide.
2. *Slide* : digunakan untuk mengatur dan mengelola yang berhubungan dengan slide. Seperti slide yang akan digunakan perlu dilakukan mengimpor soal dari file lain, atau merujuk pada internet sehingga harus melalui link.
3. *Clipboard* : digunakan untuk menyalin teks ataupun gambar.
4. *Text* : digunakan untuk mengatur tulisan dalam kuis yang akan dibuat, seperti menebalkan tulisan, memberi garis bawah tulisan, memiringkan tulisan , dan lainnya.
5. *Insert 2* : digunakan untuk menyisipkan gambar, video simbol atau lainnya pada kuis yang akan dibuat.
6. *Quiz* : digunakan untuk mengatur terkait kuis yang telah dibuat. Seperti pelengkap dalam kuis yaitu waktu, penyimpanan jawaban dan lainnya.

7. *Publish* : digunakan untuk Mengatur publikasi presentasi kuis. ada 2 fitur dalam menu *publish* yaitu *preview* dan *publish*. *Previe* dimanfaatkan untuk melihat kuis yang telah dibuat sebelum dipublikasikan. Sedangkan fitur *publish* untuk mengatur dalam mempublikasikan kuis yang telah dibuat menjadi beberapa pilihan secara *online* maupun *offline* seperti via email, web, komputer, video, atau lainnya.

2.6. PENGEMBANGAN ALAT EVALUASI

Penelitian pengembangan adalah penelitian yang digunakan untuk menghasilkan produk tertentu dan menguji keefektifan produk tersebut (Sugiyono, 2013). Banyak sekali model penelitian dan pengembangan. Salah satunya adalah model pengembangan yang dipilih oleh peneliti. Yakni model pengembangan 4D yang dikembangkan oleh Sivasailam Thiagarajan, Dorothy S. Semmel, dan Melvyn I. Semmel (1974). Model ini terdiri dari 4 tahap pengembangan yaitu Pendefinisian (*Define*), Perancangan (*Design*), Pengembangan (*Development*), dan penyebaran (*Disseminate*) dengan penjelasan sebagai berikut :

2.7.1 *Define*

Define merupakan Tahap awal dari model pengembangan 4D.

Menurut (Thiagarajan, 1974) “*The purpose of this stage is to stipulate and define instructional requirements. The initial phase is mainly analytical. Through analysis, we prescribe objectives and constraints for the instructional materials.*”

Tahapan *Define* ini bertujuan untuk mendefinisikan dan menetapkan syarat – syarat pembelajaran. Dalam mendefinisikan dan menetapkan diawali dengan analisis tujuan dan hambatan perangkat pembelajaran. Pada tahapan ini, terdiri dari beberapa kegiatan diantaranya adalah analisis awal akhir, analisis peserta didik, analisis konsep, analisis tugas, dan perumusan tujuan.

2.7.1.1 Analisis Awal Akhir (*Front – End Analysis*)

Menurut (Thiagarajan, 1974) “*Front-end analysis is the study of the basic problem facing the teacher trainer: to raise the performance levels of special education teachers. During this analysis the possibilities of more elegant and efficient alternatives to instruction are considered. Failing them, ksearch for relevant instructional materials already in circulation is conducted. If neither pertinent instructional alternatives or materials*

are available, then the development of instructional material is called for”.

Kegiatan analisis awal akhir bertujuan menetapkan masalah dasar dalam pengembangan perangkat pembelajaran. Dengan dilakukannya analisis awal akhir ini akan diperoleh gambaran fakta, harapan dan solusi alternative masalah dasar, sehingga memudahkan untuk menentukan dan memilih dalam pengembangan perangkat pembelajaran.

2.7.1.2 Analisis Peserta Didik (*Learner Analysis*)

Menurut (Thiagarajan, 1974) *“Learner analysis is the study of the target students special education teacher trainees. Student charactedstics relevant to the design and development of instruction are identified. The characteristics are entering competencies and background experiences; general attitude toward the instructional topic; and media, format, and language preferences.”*

Kegiatan analisis peserta didik bertujuan untuk mengetahui karakteristik peserta didik yang sesuai dengan rencana dan pengembanganyang telah direncanak. Karakteristik tersebut meliputi latar belakang dan perkembangan kognitif dari peserta didik serta keterampilan-keterampilan pribadi ataupun sosial yang sesuai dengan topik pembelajaran.

2.7.1.3 Analisis Tugas (*Task Analysis*)

Menurut (Thiagarajan, 1974) *“ Task analysis is the identifying of the main skill to be acquired by the teacher trainees and analyzing it into a set of necessary and sufficient subskills. This analysis ensures comprehensive coverage of the task in the instructional material.”*

Tujuan dari analisis tugas adalah mengidentifikasi keterampilan utama peserta didik yang harus diperoleh pendidik dan menganalisisnya menjadi satu set subskill yang memadai. Analisis ini yang akan menentukan cakupan materi dalam pemberian tugas yang komprehensif dengan materi pengajaran yang akan digunakan.

2.7.1.4 Analisis Konsep (*Concept Analysis*)

Menurut (Thiagarajan, 1974) *“Concept analysts is the identifying of the major concepts to be taught, arranging them in hierarchies, and breaking down individual concepts into critical and irrelevant attributes. This analysis helps to identify a rational set of examples and nonexamples to be portrayed in protocol development.”*

Analisis konsep adalah mengidentifikasi pokok-pokok atau konsep-konsep materi yang akan diajarkan pendidik kepada peserta didik serta menyusunnya secara sistematis. Analisis ini akan membantu dalam mengidentifikasi contoh dan bukan contoh untuk digambarkan dalam proses pengembangan.

2.7.1.5 Perumusan Tujuan (*Specifying Instructional Objective*)

Menurut (Thiagarajan, 1974) *“Specifying instructional objectives is the converting of the results of task and concept analyses into behaviorally stated objectives. This set of objectives provides the basis for test construction and instructional design. Later, it is integrated into the instructional materials for use by instructors and teacher trainees.”*

Perumusan tujuan instruksional merupakan merumuskan atau menetapkan hasil analisis konsep dan analisis tugas menjadi indikator pencapaian hasil belajar peserta didik yang akan menjadi tujuan khusus yang merupakan dasar atau patokan dalam penyusunan tes dan desain penelitian yang akan digunakan.

2.7.2 Design

Design atau perancangan merupakan tahapan kedua dari pengembangan yang bertujuan untuk merencanakan atau merancang produk yang akan dikembangkan sehingga dihasilkan draf awal atau *prototype*. Hal ini relevan dengan pendapat menurut (Thiagarajan, 1974) yang mengemukakan bahwa

“The purpose of this stage is to design prototype instructional material. This phase can begin after the set of behavioral objectives for the instructional material has been established. Selection of media and formats for the material and the production of an initial version constitute the major aspects of the design stage.”

Pada langkah ini, ada beberapa kegiatan yang akan dilakukan diantaranya adalah penyusunan tes acuan patokan, pemilihan media, pemilihan format, dan rancangan awal.

2.7.2.1 Penyusunan Tes Acuan Patokan (*Constituting criterion – referenced tests*)

Menurut (Thiagarajan, 1974) *“Constituting criterion-referenced tests is the step bridging Stage 1, Define, and the Design process. Criterion-referenced tests convert behavioral objectives into an outline for the instructional material.”*

Penyusunan tes acuan patokan merupakan langkah yang menghubungkan antara langkah pendefinisian dengan proses perancangan. Penyusunan tes acuan

patokan disusun berdasarkan tujuan khusus pembelajaran dan hasil analisis peserta didik. Penskoran hasil tes disesuaikan dengan panduan evaluasi, kunci dan pedoman penskoran setiap butir soal.

2.7.2.2 Pemilihan Media (*Media selection*)

Menurut (Thiagarajan, 1974) “*Media selection is the selection of appropriate media for the presentation of the, instructional content. This process involves matching the task and concept analyses, target trainee characteristics, production resources, and dissemination plans with various attributes of different media. Final selection identifies the most appropriate medium or combination of media for use.*”

Tujuan dari pemilihan media adalah memilih media yang sesuai dengan konten pembelajaran serta mengidentifikasi media yang relevan dengan karakteristik materi yang akan digunakan. Pemilihan media ini akan digunakan untuk membantu dalam proses pengembangan alat evaluasi.

2.7.2.3 Pemilihan Format (*Format selection*)

Menurut (Thiagarajan, 1974) “*Format selection is closely related to media selection. Later in this sourcebook, 21 different formats are identified which are suitable for designing in instructional materials for teacher training. The selection of the most appropriate format depends upon a number of factors, which are discussed.*”

Pemilihan Format sangat erat kaitannya dengan pemilihan media, karena format yang dipilih harus dicocokkan dengan media yang telah dipilih. Format tersebut yang dimaksudkan untuk mendesain atau merancang isi, pemilihan strategi, pendekatan, metode pembelajaran dan sumber belajar.

2.7.2.4 Rancangan Awal (*Initial design*)

Menurut (Thiagarajan, 1974) “*Initial design is the presenting of the essential instruction through appropriate media and in a suitable sequence. It also involves structuring various learning activities such as reading a text, interviewing special education personnel, and practicing different instructional skills by teaching peers.*”

Rancangan awal merupakan hasil yang telah dilakukan pada proses perancangan. Rancangan ini adalah seluruh rancangan media pembelajaran yang harus dikerjakan dan akan divalidasi sebelum diujicobakan. Ada beberapa aktifitas yang terlibat dalam langkah ini seperti membaca teks, mewawancarai guru mata pelajaran, dan praktek keterampilan – keterampilan pembelajaran seperti praktek mengajar.

2.7.3 Development

Menurut (Thiagarajan, 1974) *“The purpose of Stage III is to modify the prototype instructional material. Although much has been produced since the Define stage, the results must be considered an initial version of the instructional material which must be modified before it can become an effective final version. In the developmentstage, feedback is received through formative evaluation and the materials are suitably revised.”*

Development atau pengembangan merupakan tahapan yang akan menghasilkan bentuk akhir media pembelajaran yang dikembangkan melalui evaluasi atau validasi serta revisi untuk perbaikan. Tujuan dari tahap pengembangan ini adalah untuk menghasilkan media pembelajaran berupa alat evaluasi yang telah dibuat setelah melalui revisi berdasarkan saran dari para ahli atau validator dan akan diujicoba. Pada tahapan ini ada 2 langkah yang dilakukan yaitu penilaian ahli yang dilanjutkan dengan revisi, dan ujicoba pengembangan. Langkah-langkah tersebut adalah sebagai berikut :

2.7.3.1 Validasi Ahli (*Expert appraisal*)

Menurut (Thiagarajan, 1974) *“Expert appraisal is a technique for obtaining suggestions for the improvement of the material. A number of experts are asked to evaluate the material from instructional and technical points of view. On the basis of their feedback, the material is modified to make it more appropriate, effective, usable, and of high technical quality.”*

Penilaian para ahli merupakan teknik untuk memperoleh kevalidan, nilai, dan saran terhadap alat evaluasi yang dikembangkan yang mencakup media, ilustrasi, format, materi, isi dan lainnya. Berdasarkan saran dari para ahli, produk tersebut direvisi atau diperbaiki agar menjadi lebih tepat, efektif, efisien dan memiliki kualitas yang tinggi sehingga layak untuk digunakan dan diujicobakan.

2.7.3.2 Uji Coba Pengembangan (*Developmental testing*)

Menurut (Thiagarajan, 1974) *“Developmental testing involves trying out the material with actual trainees to locate sections for revision. On the basis of the response, reactions, and comments of the trainees, the material is modified. The cycle of testing, revising, and retesting is repeated until the material works consistently and effectively.”*

Tujuan dari ujicoba pengembangan adalah untuk memperoleh masukan secara langsung berupa respon, komentar, dan reaksi dari peserta didik dan

pengamat terhadap alat evaluasi yang telah disusun untuk menemukan bagian-bagian yang memang harus diperbaiki. Perbaikan dan uji coba terus dilakukan hingga diperoleh alat evaluasi yang valid, konsisten dan efektif.

2.7.4 Disseminate

Menurut (Thiagarajan, 1974) *“Instructional materials reach their final production stage when developmental testing yields consistent results and expert appraisal yields positive comments. The three steps in this stage are shown in Figure 1.5. Before disseminating the materials, a summative evaluation is undertaken. In its validation testing phase, the material is used under replicable conditions to demonstrate “who learns what under what conditions in how much time” (Markle, 1967). The material is also subjected to professional examination for objective opinions on its adequacy and relevance. The terminal stages of final packaging, diffusion, and adoption are most important although most frequently overlooked. A producer and a distributor must be selected and worked with cooperatively to package the material in an acceptable form. Special efforts are required to distribute the materials widely among trainers and trainees, and to encourage the adoption and utilization of the materials.”*

Tahapan yang terakhir dalam pengembangan adalah penyebaran. Setelah pengujian pengembangan alat evaluasi memperoleh hasil yang valid, konsisten dan menghasilkan komentar yang positif dari penilaian ahli dan subjek uji coba terbatas. Maka Produk alat evaluasi yang dikembangkan dapat dilakukan pada tahap penyebaran. Tahap penyebaran merupakan tahap penggunaan produk alat evaluasi yang telah dikembangkan pada skala yang lebih luas. Misalnya di kelas lain, di sekolah lain, atau oleh pendidik matematika yang lain.

Namun, ada beberapa penelitian terdahulu yang tidak melakukan tahap ini. Karena penelitiannya hanya sebatas pengembangan, sehingga hanya sampai pada tahap ketiga yaitu *development*. Hal ini sesuai dengan yang dijelaskan oleh (Andini, 2017) bahwa “peneliti membatasi langkah-langkah penelitian pengembangan dari empat langkah menjadi tiga langkah dikarenakan penelitian berupa pengembangan sehingga tahapannya hanya sampai tahap ketiga yaitu *develop* (pengembangan)”. Selain itu ada yang juga memang memodifikasi tahap pengembangan 4D menjadi 3D hal ini seperti yang dilakukan oleh (Noto, 2014).

2.7. PENELITIAN YANG RELEVAN

Sebelum melakukan penelitian ini, peneliti sudah menelusuri beberapa penelitian terdahulu yang berkaitan dengan penelitian ini. Berikut dikemukakan beberapa penelitian terdahulu yang relevan dengan penelitian ini :

1. Penelitian yang dilakukan (Girsang, 2019) mengembangkan tes *two-tier multiple choice* untuk mengukur kemampuan berpikir tingkat tinggi. Menunjukkan bahwa tes diagnostik pilihan ganda dua tingkat dengan soal bermodel HOTS tersebut dapat mengidentifikasi kemampuan berpikir tingkat tinggi dari peserta didik dengan rincian; Kualitas soal tes sangat baik dengan presentase validasi 91,33%, tes memiliki reliabilitas yang cukup adalah 0,587. Diskriminasi indeks pertanyaan diperoleh oleh 58% pertanyaan di kategori sangat baik, 21% pertanyaan dalam kategori baik, 21% pertanyaan dalam kategori buruk. Dan keefektifannya sebesar 79%. Perbedaan penelitian yang dilakukan oleh Multia dengan penelitian ini adalah instrumen yang digunakan masih dalam bentuk konvensional, sedangkan penelitian ini menggunakan *software* aplikasi yang menarik dengan berbantu IT. Namun memiliki persamaan yaitu menggunakan bentuk tes pilihan ganda dua tingkat atau disebut *two tier multiple choice*.
2. Penelitian yang dilakukan oleh (Cahyati, 2018) mengembangkan alat evaluasi pembelajaran matematika menggunakan *ispring suite*. Menunjukkan bahwa alat evaluasi yang dikembangkan dapat dikategorikan sangat layak. Hal ini dibuktikan dengan hasil validasi dari ahli materi diperoleh persentase akhir sebesar 87%, Hasil validasi dari ahli media diperoleh persentase sebesar 85,5%, serta hasil validasi dari ahli bahasa diperoleh persentase sebesar 84%. Selain itu, saat diuji cobakan kepada peserta didik, diperoleh respon dari peserta didik dengan persentase 87,6% termasuk dalam kategori “sangat menarik”. Persamaan penelitian ini dengan penelitian yang dilakukan oleh Agna Deka Cahyanti adalah menggunakan *software* aplikasi *ispring suite* dengan berbeda tipe. Untuk perbedaannya adalah penelitian Agna Deka Cahyanti membuat alat evaluasi secara umum dengan berbagai bentuk soal-soal objektif, sedangkan penelitian hanya membuat alat evaluasi khusus bentuk soal pilihan ganda dua tingkat.