



**KUALITAS INSTRUMEN TES UJIAN MADRASAH (UM)
MATA PELAJARAN GEOGRAFI DI MA NU AL-
HIKMAH KECAMATAN MIJEN SEMARANG
TAHUN PELAJARAN 2015/2016**

SKRIPSI

Untuk memperoleh gelar Sarjana Pendidikan

Oleh:

Saekhul Indrianto

3201409053

UNNES
UNIVERSITAS NEGERI SEMARANG

**JURUSAN GEOGRAFI
FAKULTAS ILMU SOSIAL
UNIVERSITAS NEGERI SEMARANG
2016**

PERSETUJUAN PEMBIMBING

Skripsi ini telah disetujui oleh pembimbing untuk diajukan ke sidang Panitia Ujian Skripsi Fakultas Ilmu Sosial pada:

Hari : Senin

Tanggal : 22 Agustus 2016

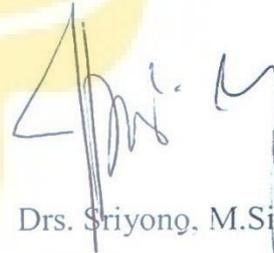
Pembimbing I



Drs. Suroso, M.Si

NIP. 196004021986011001

Pembimbing II



Drs. Sriyong, M.Si

NIP. 196312171988031002

UNNES

Mengetahui,
UNIVERSITAS NEGERI SEMARANG

Ketua Jurusan Geografi



Dr. Tjaturrahono Budi Sanjoto, M.Si.

NIP. 19621019 1988031 002

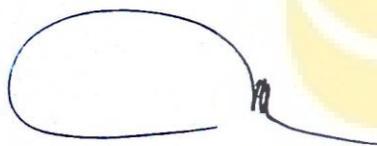
PENGESAHAN KELULUSAN

Skripsi ini telah dipertahankan di depan Sidang Panitia Ujian Skripsi
Fakultas Ilmu Sosial, Universitas Negeri Semarang pada:

Hari : *Senin*

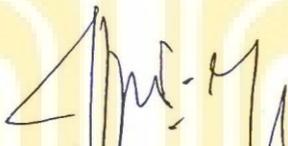
Tanggal : *22 Agustus 2016*

Penguji I



Sriyanto, S.Pd, M.Pd
NIP.19770722 2005011 001

Penguji II



Drs. Sriyono, M.Si.
NIP/ 196312171988031002

Penguji III



Drs. Suroso, M.Si
NIP. 196004021986011001

Mengetahui,

Dekan Fakultas Ilmu Sosial
Universitas Negeri Semarang



Drs. Moh Solehatul Mustofa, MA

NIP. 196308021988031001

PERNYATAAN

Penyusun menyatakan bahwa yang tertulis di dalam skripsi ini benar-benar hasil karya sendiri, bukan jiplakan dari karya tulis orang lain, baik sebagian ataupun seluruhnya. Pendapat atau temuan orang lain yang terdapat dalam skripsi ini dikutip atau dirujuk berdasarkan kode etik ilmiah.



Semarang, Agustus 2016



Saekhul Indrianto
NIM. 3201409053

MOTTO DAN PERSEMBAHAN

MOTTO

1. Allah tidak akan mengubah nasib suatu kaum apabila kaum tersebut tidak berusaha untuk mengubah nasibnya sendiri (QS. Ar Ra'du : 13).
2. Sesungguhnya bersama kesulitan ada kemudahan (QS. Al Insyirah : 6).

PERSEMBAHAN

Skripsi ini saya persembahkan untuk :

1. Kedua orang tuaku tercinta, Bapak Kariri dan Ibu Puji Wati yang selalu membangkitkanku dalam keterpurukan dan tak hentinya memberikan kasih sayang, dukungan, arahan dan do'a untuk keberhasilanku.
2. Adikku, Ari Tubety yang telah banyak memberikan dukungan kepada saya.'
3. Teman-teman Pendidikan Geografi 2009.
4. Kedua Dosen Pembimbing Bapak Suroso dan Bapak Sriyono, terimakasih atas bimbingannya yang luar biasa.
5. Almamaterku UNNES, yang telah memberikan banyak ilmu.

UNNES
UNIVERSITAS NEGERI SEMARANG

PRAKATA

Segala puji bagi Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya sehingga penyusun dapat menyelesaikan skripsi dengan judul **“Kualitas Instrumen Tes Ujian Madrasah (UM) Mata Pelajaran Geografi di MA NU Al – Hikmah Kecamatan Mijen Semarang Tahun Pelajaran 2015/2016”**.

Penulisan skripsi ini tidak lepas dari bantuan berbagai pihak yang telah banyak membantu baik motivasi, moral dan material kepada penyusun. Oleh sebab itu, dalam kesempatan ini penyusun ingin mengucapkan terimakasih kepada:

1. Prof. Dr. Fathur Rokhman, M. Hum. Rektor Universitas Negeri Semarang yang telah memberikan kesempatan untuk menyelesaikan studi strata satu di Universitas Negeri Semarang.
2. Drs. Moh Solehatul Mustofa, MA. Dekan Fakultas Ilmu Sosial Universitas Negeri Semarang yang telah memberikan kemudahan administrasi dalam perijinan penelitian.
3. Dr. Tjaturahono Budi Sanjoto, M. Si. Ketua Jurusan Geografi Fakultas Ilmu Sosial Universitas Negeri Semarang yang telah memberikan kemudahan administrasi dalam menyelesaikan skripsi ini.
4. Drs. Suroso, M.Si. Dosen Pembimbing I dan Dosen Wali yang telah memberikan bimbingan, arahan, dan saran kepada penyusun selama penyusunan skripsi ini.
5. Drs. Sriyono, M.Si. Dosen Pembimbing II yang telah memberikan bimbingan, arahan, dan saran kepada penyusun selama penyusunan skripsi ini.
6. Seluruh dosen Fakultas Ilmu Sosial Universitas Negeri Semarang khususnya dosen Jurusan Pendidikan Geografi yang telah memberikan ilmu yang tidak ternilai harganya dan mudah-mudahan dapat menjadi ilmu yang bermanfaat bagi penyusun.
7. H. A. Jumarno, S.Ag., M.Pd.I Kepala Madrasah Aliyah NU Al-Hikmah, yang telah memberi ijin penelitian di sekolahnya.

8. Murtadho, S.Pd. Guru Mata Pelajaran Geografi Madrasah Aliyah NU Al-Hikmah, yang telah membantu kelancaran penelitian.
9. Semua pihak yang tidak dapat kami sebutkan satu persatu yang telah membantu jalannya pelaksanaan penelitian ini sehingga penelitian ini dapat berjalan dengan lancar.

Hanya ucapan terima kasih dan doa, semoga apa yang telah diberikan menjadi amal baik dan mendapat balasan yang lebih baik dari Allah SWT. Penulis berharap semoga skripsi ini dapat bermanfaat bagi semua pembaca dan dapat memberikan kontribusi dalam dunia pendidikan.

Semarang, Agustus 2016

Penulis



SARI

Indrianto, Saekhul. 2016. *Kualitas Instrumen Tes Ujian Madrasah (UM) Mata Pelajaran Geografi Di MA NU Al – Hikmah Kecamatan Mijen Semarang Tahun Pelajaran 2015/2016.* Skripsi. Jurusan Geografi, Fakultas Ilmu Sosial, Universitas Negeri Semarang. Pembimbing I: Drs. Suroso, M.Si, Pembimbing II: Drs. Sriyono, M.Si. 92 halaman.

Kata Kunci: Kualitas Instrumen Tes, Ujian Madrasah (UM)

Terdapat adanya perbedaan yang cukup besar antara rata-rata nilai hasil Ujian Nasional (UN) dengan rata-rata Hasil Ujian Madrasah (UM) mata pelajaran Geografi di MA NU Al-Hikmah Kecamatan Mijen Semarang tahun pelajaran 2015/2016. Rata-rata nilai hasil Ujian Nasional (UN) mata pelajaran Geografi yaitu 54,76 sedangkan rata-rata Hasil Ujian Madrasah (UM) mata pelajaran Geografi adalah 63,08. Perbedaan yang cukup besar tersebut menunjukkan kemungkinan adanya kelemahan pada kualitas instrumen tes yang digunakan dalam Ujian Madrasah (UM). Rumusan masalah penelitian ini adalah bagaimana kualitas instrumen tes Ujian Madrasah (UM) mata pelajaran Geografi di MA NU Al-Hikmah Kecamatan Mijen Semarang tahun pelajaran 2015/2016. Tujuan penelitian ini untuk mengetahui bagaimana kualitas instrumen tes Ujian Madrasah (UM) mata pelajaran Geografi di MA NU Al-Hikmah Kecamatan Mijen Semarang tahun pelajaran 2015/2016.

Penelitian ini merupakan penelitian deskriptif dengan obyek penelitian adalah soal-soal uji Ujian Madrasah (UM) yang digunakan di MA NU Al-hikmah tahun pelajaran 2015/2016. Data diperoleh dengan menggunakan teknik dokumentasi. Data yang terkumpul dianalisis secara kuantitatif dengan menggunakan bantuan program komputer *microsoft excel 2010* untuk mengetahui validitas, reliabilitas, tingkat kesukaran, dan daya pembeda instrumen tes.

Hasil penelitian menunjukkan bahwa: (1) kualitas instrumen yang digunakan belum baik dari segi validitasnya, yaitu hanya terdapat 32% butir soal yang valid, (2) indeks reliabilitas instrumen yang digunakan adalah 0,726, hal tersebut menunjukkan bahwa reliabilitas instrumen tes yang digunakan sangat tinggi, (3) hasil uji tingkat kesukaran menunjukkan bahwa instrumen tes yang digunakan belum proporsional, (4) dilihat dari hasil analisis daya pembeda yang dilakukan, diketahui hanya terdapat 8 butir (16%) saja yang berkategori baik.

Simpulan dari penelitian ini adalah dilihat dari hasil uji validitas, reliabilitas, tingkat kesukaran, dan daya pembeda diketahui kualitas instrumen yang digunakan tergolong belum baik. Saran yang dapat peneliti berikan adalah bagi penyusun instrumen harusnya memperhatikan dengan baik prosedur penyusunan tes pilihan ganda, agar tes pilihan ganda yang dibuat dapat dipertanggungjawabkan kualitasnya. Soal yang memiliki kelemahan baik dari segi materi, konstruksi, dan bahasa yang digunakan sebaiknya diperbaiki agar

kemampuan instrumen tes untuk melihat pencapaian kompetensi siswa menjadi lebih optimal.

Penelitian ini memungkinkan untuk dilanjutkan dengan lingkup penelitian yang lebih luas. Lingkup penelitian yang dimaksud dapat berupa lingkup sekolah, banyaknya variasi soal, maupun banyaknya peserta tes.



DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
PERSETUJUAN PEMBIMBING	ii
PENGESAHAN KELULUSAN	iii
PERNYATAAN.....	iv
MOTTO DAN PERSEMBAHAN.....	v
PRAKATA	vi
SARI.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	
A. Latar Belakang	1
B. Rumusan Masalah.....	3
C. Tujuan Penelitian	3
D. Manfaat Penelitian	3
E. Penegasan Istilah	3
BAB II LANDASAN TEORI	
a. Pengertian Tes	6
b. Fungsi Tes.....	8
c. Jenis Tes.....	8
d. Prosedur Penyusunan Tes yang Baik.....	10
e. Kualitas Instrumen Tes	19
1. Validitas	20
2. Reliabilitas	25
3. Tingkat Kesukaran	30
4. Daya Pembeda	31
f. Kerangka Berfikir	32

BAB III METODE PENELITIAN

A. Populasi dan Sampel.....	34
B. Variabel Penelitian.....	35
C. Metode Pengumpulan Data	35
D. Teknik Analisis Data	36

BAB IV HASIL PENELITIAN DAN PEMBAHASAN

A. Hasil Penelitian.....	41
1. Gambaran Umum Objek Penelitian.....	41
2. Gambaran Umum Instrumen Tes Ujian Madrasah (UM)	43
3. Pelaksanaan Penelitian.....	51
4. Validitas Instrumen Tes	52
5. Reliabilitas Instrumen	53
6. Tingkat Kesukaran Instrumen Tes.....	53
7. Daya Pembeda Instrumen Tes	54
B. Pembahasan	55

BAB V PENUTUP

A. Simpulan	61
B. Saran	62

DAFTAR PUSTAKA	63
-----------------------------	----

LAMPIRAN	65
-----------------------	----



DAFTAR TABEL

Tabel	Halaman
Tabel 3.1 Kriteria Daya Pembeda (DP)	40
Tabel 4.1 Validitas Instrumen Tes	52
Tabel 4.2 Tingkat Kesukaran Instrumen Tes	53
Tabel 4.3 Indeks Daya Beda Instrumen Tes	55



DAFTAR GAMBAR

Gambar	Halaman
2.1. Kerangka Berfikir Penelitian	33
4.1. Peta Lokasi Penelitian	42



DAFTAR LAMPIRAN

Lampiran	Halaman
1. Daftar Nilai Ujian Madrasah (UM) Mata Pelajaran Geografi MA NU Al-Hikmah Tahun Pelajaran 2015/2016	65
2. Daftar Nilai Ujian Nasional (UN) Mata Pelajaran Geografi MA NU Al-Hikmah Tahun Pelajaran 2015/2016	66
3. Kisi-kisi Soal Ujian Madrasah	67
4. Analisis Soal Berdasarkan Tingkatan Kognitif (Taksonomi Bloom)	70
5. Analisis Validitas, Reliabilitas, Daya Pembeda, dan Tingkat Kesukaran Instrumen Tes	73
6. Validitas Instrumen Tes	77
7. Tingkat Kesukaran Instrumen Tes	78
8. Daya Pembeda Instrumen Tes	79
9. Perhitungan Validitas Butir Soal Nomor 1, 6, 40, 47, dan 50	80
10. Paket Soal Ujian Madrasah (UM)	82
11. Kunci Jawaban Soal Ujian Madrasah (UM)	89
12. Surat Ijin Penelitian	90
13. Surat Keterangan Penelitian	91



BAB I

PENDAHULUAN

A. Latar Belakang

Ujian Madrasah adalah kegiatan pengukuran pencapaian kompetensi peserta didik yang dilakukan oleh satuan pendidikan untuk memperoleh pengakuan atas prestasi belajar dan merupakan salah satu persyaratan kelulusan dari satuan pendidikan. Sedangkan ujian nasional (UN) sebagai kegiatan pengukuran pencapaian kompetensi peserta didik pada mata pelajaran tertentu dalam kelompok mata pelajaran ilmu pengetahuan dan teknologi dalam rangka menilai pencapaian Standar Nasional Pendidikan (Permendiknas No. 20 tahun 2007 tentang Standar Penilaian Pendidikan). Pada tahun pelajaran 2015/2016 Peserta didik dapat dinyatakan lulus dari satuan pendidikan setelah menyelesaikan seluruh program pembelajaran, memperoleh nilai sikap/perilaku minimal baik, dan lulus Ujian Sekolah/Madrasah/Pendidikan Kesetaraan (Permendikbud nomor 5 tahun 2015).

Instrumen tes yang digunakan pada Ujian Madrasah (UM) dan Ujian Nasional (UN) ditentukan berdasarkan Standar Kompetensi Lulusan (SKL) dan materi yang diujikan setiap mata pelajaran, sehingga hasil belajar yang diperoleh peserta didik baik pada Ujian Madrasah (UM) maupun Ujian Nasional (UN) seharusnya tidak jauh berbeda. Dalam Peraturan Menteri Pendidikan Nasional No. 23 Tahun 2006 tentang

Standar Kompetensi Lulusan Pasal 1 ayat 1 menyatakan bahwa: Standar Kompetensi Lulusan untuk satuan pendidikan dasar dan menengah digunakan sebagai pedoman penilaian dalam menentukan kelulusan peserta didik.

Berdasarkan dokumen yang diperoleh di MA NU Al-Hikmah, ditemukan adanya perbedaan yang cukup besar antara rata-rata nilai hasil Ujian Nasional (UN) dengan rata-rata Hasil Ujian Madrasah (UM) mata pelajaran Geografi. Rata-rata nilai hasil Ujian Nasional (UN) mata pelajaran Geografi yaitu 54,76 sedangkan rata-rata Hasil Ujian Madrasah (UM) mata pelajaran Geografi adalah 63,08. Perbedaan yang cukup besar tersebut menunjukkan kemungkinan adanya kelemahan pada kualitas instrumen tes yang digunakan dalam Ujian Madrasah (UM). Instrumen yang bermutu adalah instrumen yang dapat memberikan informasi setepat-tepatnya tentang pembelajar mana yang telah menguasai materi dan pembelajar mana yang belum menguasai materi (Kusaeri dan Suprananto, 2012: 164).

Berdasarkan latar belakang di atas, maka hal itu menjadi suatu alasan bagi penulis untuk mengangkat permasalahan tersebut dalam skripsi yang berjudul **“Kualitas Instrumen Tes Ujian Madrasah (UM) Mata Pelajaran Geografi di MA NU Al-Hikmah Kecamatan Mijen Semarang Tahun Pelajaran 2015/2016”**.

B. Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah “bagaimana kualitas instrumen tes Ujian Madrasah (UM) mata pelajaran Geografi di MA NU Al-hikmah Kecamatan Mijen Semarang tahun pelajaran 2015/2016?”.

C. Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan dari penelitian ini adalah untuk mengetahui bagaimana kualitas instrumen tes Ujian Madrasah (UM) mata pelajaran Geografi di MA NU Al-hikmah Kecamatan Mijen Semarang tahun pelajaran 2015/2016.

D. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat untuk berbagai pihak terutama dalam rangka pengembangan disiplin ilmu, peningkatan mutu pendidikan, dan untuk peneliti-peneliti lebih lanjut.

1. Bagi penulis sebagai sarana untuk menambah pengetahuan dan pengalaman terutama pada masalah yang berkaitan dengan kualitas instrumen tes.
2. Bagi sekolah/madrasah sebagai bahan penilaian dalam mengembangkan instrumen tes yang lebih baik.

E. Penegasan Istilah

Untuk menggambarkan secara jelas mengenai istilah-istilah yang merujuk pada permasalahan di atas dan sesuai dengan judul yang ingin

diteliti maka diperlukan penjelasan. Penegasan istilah bertujuan untuk menghindari segala bentuk perbedaan dalam penafsiran skripsi ini. Berbagai macam istilah yang perlu mendapatkan pembatasan adalah sebagai berikut.

1. Kualitas

Menurut Kamus Besar Bahasa Indonesia Kualitas atau mutu adalah tingkat baik buruknya atau taraf atau derajat sesuatu. Kualitas yang dimaksud dalam penelitian ini adalah kualitas instrumen tes yang digunakan pada Ujian Madrasah (UM) mata pelajaran Geografi di MANU Al-Hikmah tahun pelajaran 2015/2016 yang diketahui melalui analisis validitas, reliabilitas, tingkat kesukaran, dan daya pembeda.

2. Instrumen Tes

Instrumen merupakan sebuah alat, sedangkan tes menurut Arifin (2014:118) merupakan suatu teknik atau cara yang digunakan dalam rangka melaksanakan kegiatan pengukuran, yang di dalamnya terdapat berbagai pertanyaan, atau serangkaian tugas yang harus dikerjakan atau dijawab oleh peserta didik untuk mengukur aspek perilaku peserta didik. Sehingga instrumen tes merupakan suatu alat yang di dalamnya terdapat pertanyaan, atau serangkaian tugas yang harus dikerjakan atau dijawab oleh peserta didik untuk mengukur aspek perilaku peserta didik. Dalam penelitian ini instrumen tes yang dimaksud adalah soal-soal Ujian Madrasah (UM) mata pelajaran Geografi tahun ajaran

2015/2016 di MA NU Al-Hikmah yang berisi 50 soal berbentuk pilihan ganda.

3. Ujian Madrasah (UM)

Menurut Permendiknas No. 20 tahun 2007 tentang Standar Penilaian Pendidikan Ujian Madrasah adalah kegiatan pengukuran pencapaian kompetensi peserta didik yang dilakukan oleh satuan pendidikan untuk memperoleh pengakuan atas prestasi belajar dan merupakan salah satu persyaratan kelulusan dari satuan pendidikan. Ujian Madrasah (UM) yang dimaksud dalam penelitian ini adalah ujian yang dilaksanakan oleh Madrasah Aliyah NU Al-Hikmah Tahun Pelajaran 2015/2016 mata pelajaran Geografi yang dilaksanakan pada akhir jenjang pendidikan sebagai salah satu syarat kelulusan peserta didik.

BAB II

LANDASAN TEORI

A. Pengertian Tes

Menurut Jihad dan Haris (2013:67) tes merupakan himpunan pertanyaan yang harus dijawab, harus ditanggapi, atau tugas yang harus dilaksanakan oleh orang yang dites. Tes digunakan untuk mengukur sejauh maa seorang siswa telah menguasai pelajaran yang disampaikan terutama meliputi aspek pengetahuan dan keterampilan. Arifin (2014:118) menyatakan bahwa tes merupakan suatu teknik atau cara yang digunakan dalam rangka melaksanakan kegiatan pengukuran, yang di dalamnya terdapat berbagai pertanyaan, atau serangkaian tugas yang harus dikerjakan atau dijawab oleh peserta didik untuk mengukur aspek perilaku peserta didik.

Tes adalah suatu cara untuk melakukan penilaian yang berbentuk tugas-tugas atau pertanyaan yang harus dikerjakan oleh siswa yang sedang dites. Jawaban yang diberikan siswa terhadap pertanyaan-pertanyaan itu dianggap sebagai informasi terpercaya yang mencerminkan kemampuannya dan informasi tersebut dinyatakan sebagai masukan yang penting untuk mempertimbangkan siswa (Suwandi,2010).

Menurut Arikunto (2009:53) dalam bukunya yang berjudul *Dasar-Dasar Evaluasi Pendidikan (Edisi Revisi)* tes adalah alat atau prosedur yang digunakan untuk mengetahui atau mengukur sesuatu dalam suasana, dengan

cara dan aturan-aturan yang sudah ditentukan. Kemudian Sudijono (2006:67) menjelaskan tes adalah cara atau prosedur dalam rangka pengukuran dan penilaian di bidang pendidikan yang berbentuk pemberian tugas atau serangkaian tugas baik berupa pertanyaan yang harus dijawab atau perintah yang harus dikerjakan oleh peserta tes sehingga dapat dihasilkan nilai yang melambangkan tingkah laku atau prestasi dari peserta tes itu sendiri

Tes pada umumnya digunakan untuk mengukur proses dan hasil belajar siswa, terutama hasil belajar kognitif berkenaan dengan penguasaan bahan sesuai dengan tujuan pendidikan dan pengajaran. Meskipun begitu, dalam batas tertentu tes dapat pula digunakan untuk mengukur dan menilai hasil belajar bidang afektif dan psikomotoris (Sudjana, 2013:5).

Arikunto (2009:53) menyebutkan beberapa istilah yang terkait dengan tes, yaitu:

1. Tes: alat atau prosedur yang digunakan dalam pelaksanaan tes.
2. *Testing*: saat pada waktu pelaksanaan tes.
3. *Testee*: responden atau individu yang sedang mengerjakan tes.
4. *Tester*: orang yang melaksanakan pengambilan tes terhadap para *testee* yang bertugas untuk: (a) mempersiapkan ruangan dan perlengkapan yang diperlukan, (b) membagikan lembaran tes dan alat-alat lain untuk mengerjakan, (c) menerangkan cara mengerjakan tes, (d) mengawasi *testee* mengerjakan tes, (e) memberikan tanda-tanda waktu (f)

mengumpulkan pekerjaan *testee*, (g) Mengisi berita acara atau laporan yang diperlukan (jika ada).

B. Fungsi Tes

Beberapa fungsi tes dalam dunia pendidikan menurut Djaali (dalam Muljono 2008) adalah:

1. Sebagai alat untuk mengukur prestasi belajar siswa. Tes dimaksudkan untuk mengukur tingkat perkembangan atau kemajuan yang telah dicapai siswa setelah menempuh proses belajar-mengajar dalam jangka waktu tertentu.
2. Sebagai motivator dalam pembelajaran. Tes dianggap sebagai motivator ekstrinsik, yaitu siswa akan belajar lebih giat dan berusaha lebih keras untuk memperoleh nilai dan prestasi yang baik.
3. Sebagai upaya perbaikan kualitas pembelajaran. Dalam rangka perbaikan kualitas pembelajaran, ada tiga jenis tes yang perlu dibahas yaitu; tes penempatan, tes diagnostik, dan tes formatif.
4. Sebagai penentu berhasil atau tidaknya siswa sebagai syarat untuk melanjutkan pendidikan ke jenjang yang lebih tinggi dengan melaksanakan tes sumatif

C. Jenis Tes

Tes dapat dibedakan atas beberapa jenis, dan pembagian jenis-jenis ini dapat ditinjau dari berbagai sudut pandang. Menurut Nurgiyantoro (2011:

107), ditinjau dari segi penyusunannya, tes dibagi menjadi dua jenis, yaitu tes standar dan tes buatan guru.

1. Tes Standar

Pengertian tes standar secara sempit adalah tes yang disusun oleh tim ahli, atau disusun oleh lembaga yang menyelenggarakannya secara profesional. Tes standar juga disebut sebagai tes baku, (Arifin, 2014:120) mengemukakan bahwa tes baku adalah tes yang sudah memiliki derajat validitas dan reliabilitas yang tinggi berdasarkan percobaan-percobaan terhadap sampel yang cukup besar dan representatif. Tes baku merupakan tes yang dikaji berulang-ulang kepada sekelompok besar peserta didik, dan item-itemnya relevan serta mempunyai daya pembeda yang tinggi.

Dalam pengembangan tes standar, ciri-ciri pokok dan persyaratan tes yang baik itu dikaji secara sadar dan terencana, dan diusahakan pemenuhan kualitasnya. Penulisan tes standar biasanya dilakukan oleh sebuah tim yang sengaja dibentuk. Seleksi bahan dan tujuan didasarkan pada kurikulum atau buku teks yang dipakai secara nasional.

2. Tes Buatan Guru

Tes buatan guru adalah tes yang disusun sendiri oleh guru yang akan mempergunakan tes tersebut yang biasanya digunakan untuk

ulangan harian, formatif, dan ulangan umum (sumatif) (Arifin, 2014:119).

Kualitas tes buatan guru dilihat dari tingkat kesahihan dan keandalannya masih belum menjamin keobjektifannya, sebab hanya diberikan kepada sekelompok peserta didik, kelas, dan sekolah tertentu saja karena belum diujicobakan kepada sekelompok besar testi sehingga pengukurannya masih belum meyakinkan.

Begitu juga tingkat kesukaran itemnya tidak didasarkan atas sifat-sifat atau karakteristik peserta didiknya. Peserta didik dianggap memiliki taraf berfikir dan taraf penguasaan yang sama. Padahal setiap peserta didik secara psikologis mempunyai kemampuan yang berbeda-beda. Oleh sebab itu, sebaiknya item-item tes disusun secara cermat berdasarkan kemampuan individu yang heterogen.

D. Prosedur Penyusunan Tes yang Baik

Menurut Nurgiyantoro (2011: 20-30), kegiatan pengembangan alat penilaian perlu mengikuti beberapa langkah, yaitu penentuan spesifikasi ujian, penulisan butir soal, penelaahan butir soal, pelaksanaan uji coba, analisis butir soal dan analisis jawaban, perbaikan butir soal dan perakitan soal ujian, pelaksanaan ujian, dan penafsiran hasil ujian.

1. Penentuan spesifikasi ujian

Dalam penentuan spesifikasi ujian, ada beberapa hal yang perlu dilakukan, yaitu penentuan kompetensi dasar, pembuatan deskripsi

bahan uji, pembuatan kisi-kisi pengujian, serta penentuan bentuk soal dan lama ujian.

a. Penentuan kompetensi dasar

Ujian dimaksudkan untuk mengukur seberapa banyak pembelajar dapat mencapai kompetensi yang dibelajarkan atau yang dipelajari. Kompetensi yang diukur kadar capaiannya adalah kompetensi dasar yang dijabarkan dari standar kompetensi yang diambil dari kurikulum.

b. Pembuatan deskripsi bahan uji

Pembuatan deskripsi bahan uji yang meliputi materi pokok dan uraian materi haruslah dilakukan untuk memastikan bahan ajar apa saja yang akan diujikan. Setelah materi ditentukan selanjutnya penyusun menentukan indikator soal dan penyebaran soal.

Indikator mempunyai kontribusi yang signifikan terhadap kuantitas dan kualitas ketercapaian kompetensi. Indikator merupakan penjabaran dari kompetensi dasar (KD), yang berfungsi sebagai alat ukur dari ketercapaian kompetensi, artinya apabila peserta didik telah menguasai semua indikator berarti peserta didik tersebut telah menguasai kompetensi. Indikator yang telah dirumuskan juga merupakan indikator soal sebagai salah satu instrumen pengukuran atau penilaian dalam sistem penilaian dalam Kurikulum Tingkat Satuan Pendidikan (KTSP). Dalam

merumuskan indikator pembelajaran langkah kerja yang harus dilakukan oleh seorang pendidik adalah:

- 1) Menganalisis standar kompetensi. Apabila KD yang tersedia pada standar kompetensi tersebut belum mampu mengakomodir seluruh amanat yang terdapat pada Standar Kompetensi, guru harus menambahkan rumusan KD hingga amanat dalam Standar Kompetensi dapat diakomodir.
- 2) Menganalisis KD. Ada beberapa hal yang harus menjadi perhatian dalam menganalisis KD, antara lain: (1) Kata Kerja Operasional (KKO). KKO yang digunakan berada pada ranah kognitif, ingatan (C1), pemahaman (C2), aplikasi (C3), analisis (C4), sintesis (C5), atau evaluatif (C6). Hal ini perlu dilakukan karena KKO pada indikator tidak boleh lebih tinggi dari KKO pada KD, paling tinggi hanya sama karena indikator fungsinya menjabarkan KD. (2) Menggaris bawahi amanat yang terdapat dalam KD tersebut. (3) Menganalisis amanat yang telah digaris bawahi. Hal ini diperlukan karena apabila amanat tersebut tidak dapat dicapai dalam satu langkah perlu dirumuskan indikator perantara atau indikator penunjang.
- 3) Menganalisis materi pembelajaran. Hal ini diperlukan karena dalam memilih dan menetapkan materi ada beberapa aspek yang harus dipertimbangkan, yaitu: (1) kontekstual, artinya materi tersebut harus mempunyai korelasi dengan keseharian

peserta didik, (2) Visi dan misi sekolah, artinya bahwa materi yang ditetapkan memiliki titik singgung dengan visi sekolah, (3) perluasan dan pengembangan materi. Ketiga aspek ini tentu memerlukan evaluasi untuk itu perlu dirumuskan indikator yang berkaitan dengan masalah tersebut.

- 4) Merumuskan indikator dengan menggunakan bahasa Indonesia yang baik dan benar, simpel, jelas, dan mudah dipahami.

Pengembangan indikator hendaknya memperhatikan UKRK (urgensi, kontinuitas, relevansi, dan keterpakaian). Urgensi, maksudnya penting dan harus dikuasai oleh peserta didik. Kontinuitas, yaitu pendalaman dan/atau perluasan dari kompetensi pada tingkat/jenjang sebelumnya. Relevansi, diperlukan karena ada hubungannya untuk mempelajari atau memahami kompetensi dan/atau konsep mata pelajaran lain. Keterpakaian, artinya memiliki nilai terapan tinggi dalam kehidupan sehari-hari. Selain itu rumusan indikator yang baik harus memenuhi beberapa persyaratan antara lain adalah:

- 1) Dirumuskan dalam kalimat yang simpel, jelas, dan mudah dipahami.
- 2) Menggunakan bahasa Indonesia yang baik dan benar.
- 3) Tidak menggunakan makna yang bermakna ganda.
- 4) Hanya mengandung satu tindakan.
- 5) Menggunakan kata kerja operasional (KKO) yang dapat diukur.

- 6) Menggunakan KKO yang lebih rendah tingkatannya dan atau sama dengan jumlah amanat yang terdapat pada KD.
- 7) Jumlah indikator minimal untuk satu KD sama dengan jumlah amanat yang terdapat pada KD tersebut.
- 8) Dalam satu KD harus ada indikator yang mengacu sekurangnya pada 2 dari 3 aspek kompetensi (*cognitive, affectif, dan psychomotor*)

Seorang pendidik dalam merumuskan suatu alat ukur berupa tes harus dapat menyesuaikan bobot soal dengan tuntutan kompetensi yang harus dimiliki peserta didik selama mengikuti pembelajaran. Bobot soal yang dimaksudkan yaitu berupa tingkatan kompetensi berdasarkan Taksonomi Bloom mulai dari tingkat pengetahuan sampai evaluasi. Jadi tiap penyusunan butir soal disesuaikan dengan tingkatan hasil belajar yang harus dikuasai oleh peserta didik sebagai bentuk pencapaian kompetensi yang sesuai dengan silabus pembelajaran.

c. Pembuatan kisi-kisi pengujian

Pengembangan alat pengujian harus mengukur semua kompetensi dasar (yang tercermin dalam bahan ajar dan indikator) secara proporsional terhadap semua kompetensi dasar yang diujikan, sehingga pembuatan butir-butir soal harus mendasarkan diri pada kisi-kisi yang sengaja dirancang sebagai maksud agar

kegiatan pengembangan butir soal dapat dilakukan dengan baik dan dapat dipertanggungjawabkan.

d. Penentuan bentuk soal dan lama ujian

Bentuk soal yang akan dipilih harus direncanakan juga, misalnya apakah bentuk objektif dengan segala sub jenisnya. Selain itu, untuk menentukan berapa jumlah butir soal yang akan diujikan, harus pula memperhitungkan waktu yang tersedia.

2. Penulisan butir soal

Penulisan butir soal adalah membuat tagihan sesuai dengan tuntutan indikator dan yang sesuai pula dengan bahan ajar. Jadi, butir-butir soal haruslah sesuai dengan bunyi “tuntutan” indikator yang bersangkutan sebagaimana tercermin pada kata kerja operasionalnya.

3. Penelaahan butir soal

Sebagai sebuah alat ukur keberhasilan pembelajaran, agar kualitasnya dapat dipertanggungjawabkan, soal harus memenuhi persyaratan sebagai alat tes yang baik. Butir soal yang telah memenuhi kriteria dari aspek materi, konstruksi, dan bahasa akan dinyatakan sudah memenuhi validitas isi atau validitas secara rasional. Surapranata (2009: 50) menambahkan bahwa telaah soal ini kemudian disebut juga analisis butir soal secara rasional atau analisis kualitatif butir soal, oleh karena soal-soal yang telah disusun tersebut kemudian ditelaah dengan menggunakan pertimbangan rasio/logika keilmuan. Dengan adanya telaah butir soal ini akan ditemukan berbagai kesalahan atau kekeliruan

yang dapat mengganggu, dan sebaliknya, juga akan dapat dipastikan kualitas butir soal yang bersangkutan. Pada instrumen tes bentuk pilihan ganda kaidah penulisan soalnya adalah sebagai berikut (Kemendikbud, 2015:16):

a. Substansi/Materi

- 1) Soal sesuai dengan indikator (menuntut tes bentuk PG).
- 2) Materi yang diukur sesuai dengan kompetensi (UKRK: urgensi, keberlanjutan, relevansi, dan keterpakaian).
- 3) Pilihan jawaban homogen dan logis.
- 4) Hanya ada satu kunci jawaban yang tepat.

b. Konstruksi

- 1) Pokok soal dirumuskan dengan singkat, jelas, dan tegas.
- 2) Rumusan pokok soal dan pilihan jawaban merupakan pernyataan yang diperlukan saja.
- 3) Pokok soal tidak memberi petunjuk kunci jawaban.
- 4) Pokok soal tidak menggunakan pernyataan negatif ganda.
- 5) Gambar/grafik/tabel/diagram dan sebagainya jelas dan berfungsi.
- 6) Panjang rumusan pilihan jawaban relatif sama.
- 7) Pilihan jawaban tidak menggunakan pernyataan "semua jawaban benar" atau "semua jawaban salah".
- 8) Pilihan jawaban yang berbentuk angka atau waktu disusun berdasarkan besar kecilnya angka atau kronologis kejadian.

9) Butir soal tidak bergantung pada jawaban soal sebelumnya.

c. Bahasa

- 1) Menggunakan bahasa sesuai dengan kaidah Bahasa Indonesia.
- 2) Menggunakan bahasa yang komunikatif.
- 3) Pilihan jawaban tidak mengulang kata/kelompok kata yang sama, kecuali merupakan satu kesatuan pengertian.
- 4) Tidak menggunakan bahasa yang berlaku setempat/tabu.

4. Pelaksanaan uji coba

Uji coba alat evaluasi dilaksanakan di kelas setelah semua persyaratan penyusunan butir soal sudah terpenuhi. Dari pelaksanaan pengukuran inilah akan diperoleh data empiris yang menunjukkan kualitas atau informasi tentang alat tes yang bersangkutan.

5. Analisis butir soal dan analisis jawaban

a. Analisis butir soal

Setelah soal selesai diujicobakan, maka soal tersebut harus dianalisis secara empiris, dalam artian bahwa butir-butir soal tersebut dianalisis berdasarkan sifat statistiknya, sehingga analisis ini disebut juga analisis butir soal secara kuantitatif.

b. Analisis jawaban

Analisis jawaban dimaksudkan untuk mengetahui kemampuan dasar/indikator mana saja yang sudah dikuasai pembelajar dan mana yang belum. Berdasarkan hasil telaah itu kemudian dapat ditentukan tindak lanjut yang perlu diambil,

misalnya apakah perlu adanya program remedial, penguatan/pengayaan, atau yang lain.

6. Perbaikan butir soal dan perakitan soal ujian

Butir-butir soal yang sudah dianalisis kemudian diperbaiki, butir-butir soal tersebut lalu disempurnakan untuk dijadikan sebuah perangkat tes yang siap untuk dipergunakan.

7. Pelaksanaan ujian

Pelaksanaan dalam proses pengukuran ini harus dilakukan sebaik mungkin dengan pengawasan yang cermat, tetapi tidak mengganggu peserta tes. Misalnya, menjalani ujian dalam kondisi kejiwaan tertekan akan memberikan hasil yang tidak sama dengan ujian dalam kondisi tidak tertekan.

8. Penafsiran hasil ujian

Pelaksanaan pengujian akan menghasilkan data empiris kuantitatif yang berwujud skor-skor untuk tiap pembelajar. Sejalan dengan prinsip keseluruhan proses penilaian, skor-skor tersebut kemudian ditafsirkan untuk memberikan makna capaian pembelajar. Pertama, pemberian makna untuk menentukan nilai seorang pembelajar yang lazimnya diberikan dalam wujud angka atau huruf (transformasi dari skala angka). Kedua, skor seorang pembelajar dapat juga dimaknai sebagai seberapa banyak ia dapat menyerap, menguasai, atau melakukan berbagai kompetensi yang dibelajarkan. Ketiga, tinggi rendahnya

capaian pembelajar juga dapat diartikan sebagai keberhasilan/kegagalan guru dalam membelajarkan mereka.

E. Kualitas Instrumen Tes

Dalam penilaian hasil belajar, tes diharapkan dapat menggambarkan sampel perilaku dan dapat menghasilkan nilai yang objektif serta akurat. Jika tes yang digunakan guru kurang baik, maka hasil yang diperoleh pun tentunya kurang baik. Hal ini dapat merugikan peserta didik itu sendiri. Artinya hasil yang diperoleh peserta didik menjadi tidak objektif dan tidak adil. Oleh sebab itu, tes yang digunakan guru haruslah memiliki kualitas yang baik dilihat dari berbagai segi. Sudjana (2013:12) menyatakan bahwa keberhasilan mengungkapkan hasil dan proses belajar siswa sebagaimana adanya (objektivitas hasil penilaian) sangat bergantung pada kualitas instrumen penilaiannya di samping pada cara pelaksanaannya.

Menurut Sudjono (2011: 369), salah satu cara untuk mengantisipasi permasalahan mengenai kecenderungan kesalahan penafsiran hasil tes buatan guru yang diakibatkan oleh alat ukur yang tidak berkualitas adalah dengan jalan melakukan penganalisisan terhadap tes yang telah dijadikan alat ukur dalam rangka mengukur keberhasilan belajar dari para peserta tes tersebut. Penelusuran atau pelacakan tersebut dilakukan oleh guru dengan tujuan untuk mengetahui, apakah butir-butir soal yang membangun tes itu sudah dapat menjalankan fungsinya sebagai alat pengukur hasil belajar yang memadai atau belum. Arifin (2014: 246) juga menyatakan bahwa analisis butir soal merupakan suatu tahap yang harus ditempuh untuk

mengetahui derajat kualitas suatu tes, baik tes secara keseluruhan maupun butir soal yang menjadi bagian dari tes tersebut.

Melalui analisis butir soal guru dapat memperoleh informasi mengenai kualitas instrumen yang digunakan. Seperti yang dijelaskan oleh Arikunto (2012: 207) yang berpendapat bahwa dengan analisis butir soal, akan diperoleh informasi tentang kekurangan sebuah soal dan petunjuk untuk mengadakan perbaikan. Kusaeri dan Suprananto (2012: 164) juga berpendapat bahwa analisis butir soal memberikan manfaat: (1) menentukan soal-soal yang cacat atau tidak berfungsi dengan baik, (2) meningkatkan butir soal melalui tiga komponen analisis yakni tingkat kesukaran, daya pembeda, dan pengecoh soal, serta (3) merevisi butir soal yang tidak relevan dengan materi yang diajarkan, ditandai dengan banyaknya peserta tes yang tidak dapat menjawab butir soal tertentu.

Jihad dan Haris (2013:179) menyatakan bahwa kriteria instrumen tes yang berkualitas dapat dilihat dari validitas dan reliabilitas instrumen tes, serta tingkat kesukaran dan daya pembeda butir soal. Analisis kualitas soal pada penelitian ini merujuk pada teori tes klasik, yaitu proses penelaahan butir soal melalui informasi dari jawaban peserta didik guna meningkatkan mutu butir soal yang bersangkutan.

1. Validitas

Sudijono (2006) menyatakan bahwa validitas yaitu ketepatan mengukur yang dimiliki oleh sebutir item (yang merupakan bagian tak terpisahkan dari tes sebagai suatu totalitas), dalam mengukur apa yang

seharusnya diukur lewat butir item tersebut. Mengacu Djaali dalam Sujarwadi (2011), validitas butir tercermin pada besaran koefisien korelasi antara skor butir dengan skor total instrumen. Untuk menghitung koefisien korelasi antara skor butir dengan skor total instrumen, digunakan rumus statistika yang sesuai dengan jenis skor butir dari instrumen tersebut. Jika diketahui bahwa skor butir kontinum, maka dapat digunakan rumus korelasi *product moment*. Jika skor butir dikotomi (misalnya 0,1) maka digunakan rumus koefisien korelasi biserial.

Nilai koefisien korelasi yang didapat dibandingkan dengan nilai koefisien korelasi yang ada di tabel r pada alpha tertentu. Jika koefisien korelasi hitung lebih besar dari tabel r, maka koefisien korelasi butir signifikan dan butir tersebut valid. Suatu teknik evaluasi dikatakan mempunyai validitas yang tinggi (disebut valid) jika teknik atau tes itu dapat mengukur apa yang sebenarnya akan diukur (Purwanto, 2009:138).

Validitas berasal dari kata *validity* yang mempunyai arti sejauh mana ketepatan dan kecermatan suatu instrument pengukuran (tes) dalam melakukan fungsi ukurnya. Suatu tes dapat dikatakan mempunyai validitas yang tinggi apabila tes tersebut menjalankan fungsi ukurnya (Azwar, 2007:174). Validitas dibagi menjadi tiga, yaitu validitas isi, validitas empiris dan validitas konstruk.

a. Validitas Isi

Validitas Isi adalah validitas yang dilihat dari segi isi tes itu sendiri sebagai alat pengukur hasil belajar yaitu: sejauh mana tes hasil belajar sebagai alat ukur hasil belajar peserta didik, isinya telah dapat mewakili secara representatif terhadap keseluruhan materi atau bahan pelajaran yang seharusnya ditekankan (Sudijono, 2009). Jadi dengan adanya analisis tes mengenai kevalidan isi terdapat ketelitian dan kesesuaian dari semua aspek yang telah dirumuskan dalam kisi-kisi dan Rencana Pelaksanaan Pembelajaran (RPP).

b. Validitas Empiris

Validitas Empiris adalah cara analisis suatu tes dengan menggunakan teknik statistik, yaitu analisis korelasi. Validitas empiris dapat diketahui berdasarkan pengamatan di lapangan. Jadi validitas empiris dapat diketahui berdasarkan ketepatan mengukur hasil belajar peserta didik yang hendak diukur dengan tes tersebut. Validitas empiris dilakukan dari dua segi yaitu dari segi ketepatan meramalnya disebut validitas ramalan (*Predictive Validity*) dan dari segi ketepatan bandingannya disebut validitas bandingan (*Concurrent Validity*).

c. Validitas Konstruk

Validitas Konstruk berkenaan dengan kesanggupan alat penilaian untuk mengukur pengertian-pengertian yang terkandung dalam materi yang diukurnya (Sudjana, 2004). Validitas konstruksi dari suatu tes hasil belajar dapat dilakukan penganalisisannya dengan jalan melakukan pencocokan antara aspek-aspek berpikir yang terkandung dalam tes hasil belajar tersebut, dengan aspek berpikir yang dikehendaki untuk diungkap oleh tujuan instruksional khusus. Tujuan instruksional pada hakikatnya adalah perubahan tingkahlaku yang diinginkan pada diri peserta didik (Sudjana, 2004).

Validitas suatu instrumen dapat diketahui dengan cara menganalisis suatu tes dengan menggunakan teknik statistik, yaitu analisis korelasi. Cara ini digunakan untuk mengetahui validitas empiris suatu instrumen. Validitas empiris dapat diketahui berdasarkan pengamatan di lapangan. Jadi validitas empiris dapat diketahui berdasarkan ketepatan mengukur hasil belajar peserta didik yang hendak diukur dengan tes tersebut.

Banyak faktor yang dapat mempengaruhi hasil tes tidak valid. Beberapa faktor tersebut secara garis besar dapat dibedakan menurut sumbernya, yaitu:

a. Faktor yang berasal dari dalam tes.

Beberapa sumber yang pada umumnya berasal dari faktor internal tes evaluasi di antaranya sebagai berikut.

- 1) Arahan tes yang disusun dengan makna tidak jelas sehingga dapat mengurangi validitas tes.
- 2) Kata-kata yang digunakan dalam struktur instrument evaluasi, terlalu sulit.
- 3) Item-item tes dikonstruksi dengan jelek.
- 4) Tingkat kesulitan item tes tidak tepat dengan materi pembelajaran yang diterima siswa.
- 5) Waktu yang dialokasikan tidak tepat, hal ini termasuk kemungkinan terlalu kurang atau terlalu longgar.
- 6) Jumlah item tes terlalu sedikit sehingga tidak mewakili sampel materi pembelajaran.
- 7) Jawaban masing-masing item evaluasi bisa diprediksi siswa.

b. Faktor yang berasal dari administrasi dan skor.

- 1) Waktu pengerjaan tidak cukup sehingga siswa dalam memberikan jawaban dalam situasi yang tergesa-gesa.
- 2) Adanya kecurangan dalam tes sehingga tidak bisa membedakan antara siswa yang belajar dengan yang melakukan kecurangan.
- 3) Pemberian petunjuk dari pengawas yang tidak dapat dilakukan pada semua siswa.

- 4) Teknik pemberian skor yang tidak konsisten, misalnya pada tes esai, juga dapat mengurangi validitas tes evaluasi.
 - 5) Siswa tidak dapat mengikuti arahan yang diberikan dalam tes esai, juga dapat mengurangi validitas tes evaluasi.
 - 6) Siswa tidak dapat mengikuti arahan yang diberikan dalam tes baku.
- c. Faktor-faktor yang berasal dari jawaban siswa.

Seringkali terjadi bahwa interpretasi terhadap item-item tes evaluasi tidak valid, karena dipengaruhi oleh jawaban siswa dari pada interpretasi item-item pada tes evaluasi.

2. Reliabilitas

Arifin (2014:258) dalam bukunya yang berjudul *Evaluasi Pembelajaran* menjelaskan, reliabilitas adalah tingkat atau derajat konsistensi dari suatu instrumen. Suatu alat tes dikatakan reliabel apabila alat tes tersebut selalu memberikan hasil yang sama apabila alat tes tersebut diteskan atau diujikan pada kelompok yang sama pada waktu atau kesempatan yang berbeda. Menurut Azwar (2007), secara tradisional terdapat tiga macam pendekatan reliabilitas yaitu :

- a. Pendekatan Tes Ulang (*tes-retest*)

Pendekatan ini menunjukkan konsistensi pengukuran dari waktu ke waktu dan menghasilkan koefisien reliabilitas yang dinamakan koefisien stabilitas. Prinsip estimainya adalah menyajikan tes dua kali pada satu kelompok subjek dengan

tenggang waktu tertentu. Asumsinya adalah bahwa suatu tes yang reliabel tentu akan menghasilkan skor tampak yang relative sama apabila dikenakan dua kali pada waktu yang berbeda. Semakin besar variasi perbedaan skor subjek antara kedua pengenaan tes, berarti semakin sulit untuk mempercayai bahwa tes itu memberikan hasil ukur yang konsisten (Azwar, 2007).

Dalam prakteknya, pendekatan ini memiliki keterbatasan yaitu kurang praktis dalam pelaksanaan tes sebanyak dua kali, dan besarnya kemungkinan terjadi efek bawaan (*carry-over effects*) dari satu pengenaan tes ke pengenaan tes yang kedua. Artinya dalam hal ini besar kemungkinan terjadinya proses pembelajaran bagi subjek dan hal ini akan mempengaruhi terhadap skor hasil tes yang kedua kalinya.

b. Pendekatan Bentuk-Paralel (*Alternate-forms*)

Pendekatan bentuk paralel merupakan pengenaan dua bentuk tes paralel dalam waktu yang bersamaan pada satu kelompok subjek. Jadi pendekatan ini hanya dapat dilakukan apabila tersedia dua bentuk instrumen yang dapat dianggap memenuhi asumsi paralel. Salah satu indikator terpenuhinya asumsi paralel adalah setaranya korelasi antara skor kedua instrument tersebut dengan skor suatu ukuran lain. Tentu saja untuk dapat paralel kedua bentuk instrumen harus disusun dengan tujuan mengukur objek

psikologis yang sama, berdasarkan *blue-print* (pola-rancangan) yang sama dan spesifikasi yang sama pula.

Kelemahan utama dari pendekatan ini adalah sulitnya menyusun dua alat ukur yang memenuhi persyaratan parallel atau sejajar. Selain kelemahan di atas, pendekatan ini juga memiliki kelebihan, yaitu dapat menghilangkan kemungkinan terjadinya efek bawaan.

c. Pendekatan Konsistensi Internal (*Internal-Consistency*)

Pendekatan konsistensi internal dilakukan dengan menggunakan satu bentuk tes yang dikenakan hanya sekali saja pada sekelompok subjek (*single-trial administration*). Pendekatan ini mengatasi kelemahan pada pendekatan *tes-retest* dan *alternate-forms*.

Pendekatan reliabilitas konsistensi internal bertujuan untuk melihat konsistensi antar-item atau antar bagian dalam tes itu sendiri. Untuk itu, setelah skor setiap item diperoleh dari sekelompok subjek, tes dibagi menjadi beberapa belahan, bisa dua belahan, tiga belahan dan bahkan belahan sebanyak item.

Membelah tes prinsipnya adalah mengusahakan agar antar belahan memiliki jumlah item sama banyak, taraf kesulitan seimbang, isi sebanding, dan memenuhi ciri-ciri paralel. Bentuk dan sifat alat ukur serta banyaknya belahan yang dibuat akan menentukan teknik perhitungan koefisien reliabilitasnya. Ada

beberapa teknik komputasi reliabilitas konsistensi internal diantaranya: formula *Spearman-Brown*, formula *Rulon*, Formula *Alpha*, formula-formula *Kuder-Richardson*, Formula *Kristof*, Formula Analisis *varians* dan sebagainya.

Indeks reliabilitas alat penilaian dapat dicari dengan mengorelasikan skor-skor yang diperoleh dari hasil penilaian yang berulang-ulang pada waktu yang berbeda atau dengan kelompok pertanyaan yang sepadan (Sudjana, 2013:16). Untuk menghitung reliabilitas tanpa melakukan korelasi dari dua pengukuran digunakan cara yang disebut dengan kesamaan rasional. Cara ini dilakukan dengan menghubungkan setiap butir soal dalam satu tes dengan butir-butir lainnya dalam tes itu sendiri secara keseluruhan.

Teknik *Kuder-Richardson* biasanya digunakan jika suatu instrumen mengukur satu gejala psikologis atau perilaku yang sama. Artinya tes tersebut dapat dikatakan reliabel bila terbukti ada konsistensi jawaban antara soal yang satu dengan soal yang lain.

Crocker & Algina (2005) menjelaskan bahwa ada 3 hal utama yang secara tidak langsung mempengaruhi tinggi rendahnya koefisien reliabilitas suatu instrumen, yaitu:

a. Homogenitas Kelompok

Secara umum dapat dijelaskan bahwa dalam suatu kondisi tes, semakin besar homogenitas kelompok berkaitan dengan trait-trait tertentu yang diukur maka indeks reliabilitas akan semakin rendah

bila dibandingkan dengan kondisi ketika kelompok sampel lebih heterogen.

b. Batasan Waktu dalam Tes

Tes yang memiliki waktu yang lebih panjang cenderung akan memiliki reliabilitas yang lebih tinggi dibandingkan tes yang memiliki waktu yang lebih pendek, terutama pada tes dengan komposisi item yang sama. Hal ini dikarenakan performansi subjek pada tes yang lebih panjang waktunya akan lebih maksimal. Sementara pada tes yang memiliki waktu lebih pendek, performansi subjek akan sangat ditentukan oleh banyak faktor, termasuk kelelahan dan performansi subjek lain yang mengikuti tes tersebut.

c. Panjang Tes

Panjang dari suatu tes sangat bergantung dengan seberapa banyaknya item-item yang menyusun tes tersebut. Secara umum, semakin banyak item yang memiliki kualitas baik dalam suatu tes, maka semakin tinggi pula indeks reliabilitas instrumen tersebut.

Faktor-faktor lain yang juga mempengaruhi reliabilitas instrument tes di antaranya sebagai berikut:

- a. Panjang tes, semakin panjang suatu tes evaluasi, semakin banyak jumlah item materi pembelajaran diukur.
- b. Penyebaran skor, koefisien reliabilitas secara langsung dipengaruhi oleh bentuk sebaran skor dalam kelompok siswa yang diukur.

- c. Kesulitan tes, tes normatif yang terlalu mudah atau terlalu sulit untuk siswa cenderung menghasilkan skor reliabilitas rendah.
- d. Obyektivitas, yang dimaksud dengan obyektif yaitu derajat di mana siswa dengan kompetensi sama, mencapai hasil sama (Sukardi, 2008:52).

Sedangkan menurut Ngalim Purwanto (2002:141), faktor-faktor yang dapat mempengaruhi keandalan suatu tes adalah:

- a. Luas-tidaknya sampling yang diambil. Makin luas suatu sampling, berarti tes semakin handal.
- b. Perbedaan bakat dan kemampuan murid yang dites. Makin variabel kemampuan peserta tes, berarti makin tinggi keadaan koefisien tes.
- c. Suasana dan kondisi testing. Suasana ketika berlangsung *testing*, seperti tenang, gaduh, banyak gangguan, pengetes yang marah-marah dapat mengganggu mengerjakan tes sehingga dengan demikian mempengaruhi pula hasil dan keandalan tes.

3. Tingkat Kesukaran

Tingkat kesukaran soal adalah peluang untuk menjawab benar suatu soal pada tingkat kemampuan tertentu yang biasa dinyatakan dengan indeks (Arifin, 2009). Kualitas instrumen tes yang baik, di samping memenuhi validitas dan reliabilitas adalah daya keseimbangan dari tingkat kesukaran soal tersebut. Keseimbangan yang dimaksudkan adalah adanya soal-soal yang termasuk mudah sedang dan sukar secara

proporsional. Tingkat kesukaran soal dipandang dari kesanggupan atau kemampuan siswa dalam menjawabnya, bukan dilihat dari segi guru dalam melakukan analisis pembuat soal (Sudjana, 2013:135). Sejalan dengan pernyataan di atas, Arifien (2014:266) juga menyatakan jika suatu soal memiliki tingkat kesukaran seimbang (proporsional), maka dapat dikatakan bahwa soal tersebut baik. Satu soal tes hendaknya tidak terlalu sukar dan tidak terlalu mudah.

Untuk memperoleh prestasi belajar yang baik, sebaiknya proporsi antara tingkat kesukaran soal tersebar secara normal. Perhitungan proporsi tersebut dapat diatur sebagai berikut:

- a. Soal sukar 25%, soal sedang 50%, soal mudah 25%, atau
- b. Soal sukar 20%, soal sedang 60%, soal mudah 20%, atau
- c. Soal sukar 15%, soal sedang 70%, soal mudah 15%.

4. Daya Pembeda

Analisis daya pembeda mengkaji butir-butir soal dengan tujuan untuk mengetahui kesanggupan soal dalam membedakan siswa yang tergolong mampu (tinggi prestasinya) dengan siswa yang tergolong kurang atau lemah prestasinya. Artinya, bila soal diberikan kepada anak yang mampu, maka hasilnya akan menunjukkan prestasi yang tinggi; dan bila diberikan kepada siswa yang lemah, hasilnya akan menunjukkan prestasi yang rendah (Sudjana, 2013:141). Dari pernyataan tersebut dapat disimpulkan bahwa daya pembeda adalah kemampuan suatu butir

tes untuk dapat membedakan (mendiskriminasi) antara siswa yang berkemampuan tinggi dengan siswa yang kemampuannya rendah.

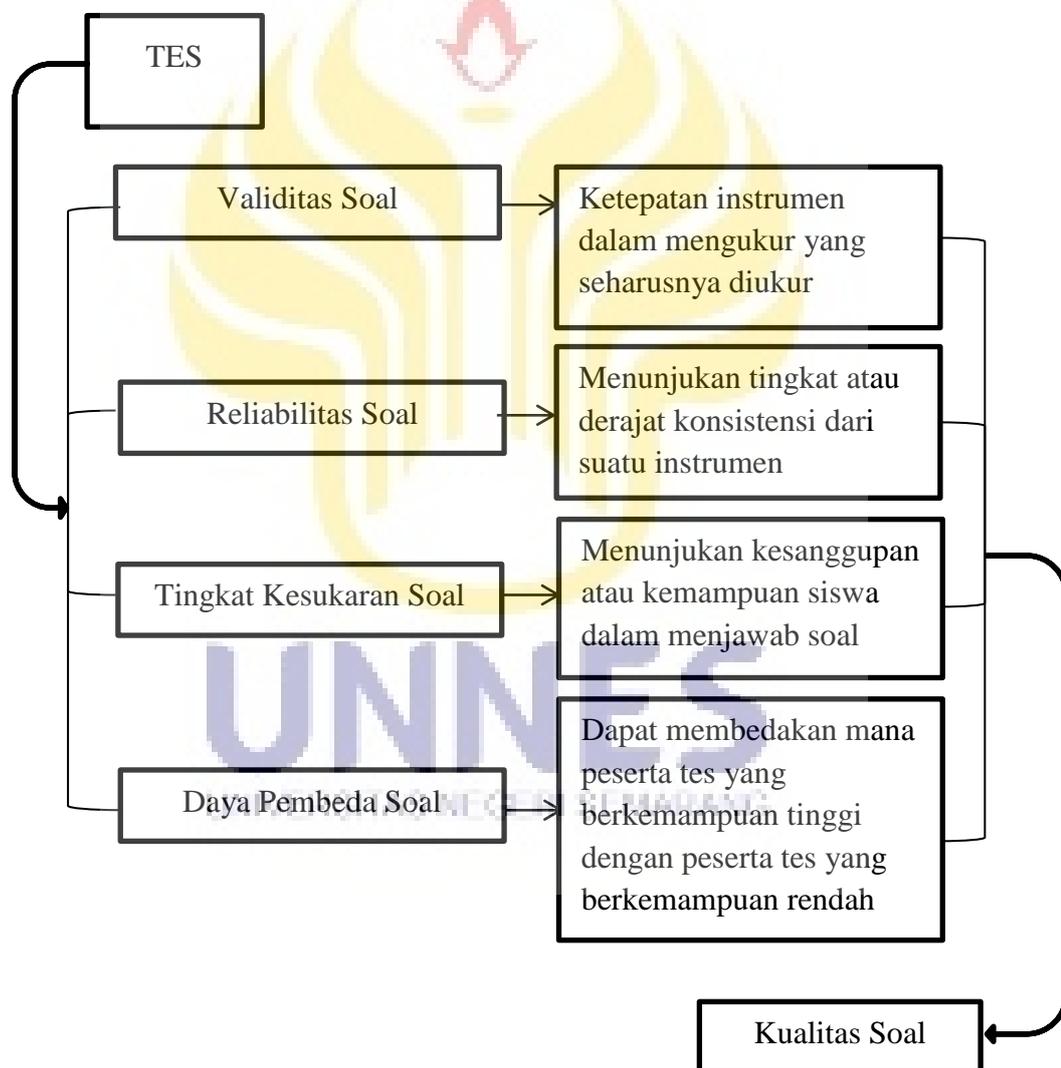
Daya pembeda soal dapat diketahui dengan melihat besar kecilnya angka indeks diskriminasi soal. Angka indeks diskriminasi soal adalah sebuah angka atau bilangan yang menunjukkan besar kecilnya daya pembeda yang dimiliki oleh butir soal. Bagi soal yang dapat dijawab benar oleh peserta didik yang menguasai materi maupun yang belum menguasai materi, maka soal itu tidak baik karena tidak mempunyai daya pembeda. Demikian pula jika semua siswa baik pandai maupun bodoh tidak dapat menjawab dengan benar soal tersebut tidak baik juga, karena tidak mempunyai daya pembeda. Soal yang baik adalah soal yang dapat dijawab benar oleh peserta didik yang pandai (Arikunto, 2012:227).

F. Kerangka Berfikir

Ujian Madrasah merupakan kegiatan pengukuran pencapaian kompetensi peserta didik yang dilakukan oleh Madrasah dan merupakan salah satu persyaratan kelulusan dari Madrasah. Sebagai syarat kelulusan maka soal yang dipakai pada ujian tersebut harus memenuhi persyaratan tes yang baik sehingga dapat dengan tepat mengukur kemampuan siswa. Dengan kata lain tes yang digunakan haruslah berkualitas.

Untuk mengetahui kualitas suatu instrumen tes Ujian Madrasah (UM) tersebut, dapat dilakukan analisis per butir soal menggunakan jawaban siswa. Analisis tersebut meliputi analisis validitas, reliabilitas,

tingkat kesukaran, daya pembeda. Hasil analisis yang diperoleh kemudian digunakan untuk menentukan seberapa baik kualitas soal Ujian Madrasah yang digunakan di MA NU Al-Hikmah. berikut ini merupakan kerangka berfikir penelitian dalam bentuk bagan.



Gambar 2.1. Kerangka Berfikir Penelitian

BAB V

SIMPULAN DAN SARAN

A. Simpulan

Berdasarkan hasil penelitian serta pembahasan yang terdapat pada bab sebelumnya, maka dapat disimpulkan sebagai berikut:

1. Hasil uji validitas butir soal yang dilakukan menunjukkan bahwa kualitas instrumen tes yang digunakan masih belum baik, dimana diketahui hanya terdapat 16 (32%) dari 50 butir saja yang dinyatakan valid.
2. Instrumen tes yang dipakai berdasarkan hasil uji reliabilitas dinyatakan memiliki kualitas yang sangat baik, hal tersebut dapat dilihat dari besarnya indeks reliabilitas yaitu 0,726 (reliabilitas sangat tinggi).
3. Proporsi tingkat kesukaran pada instrumen tes yang digunakan belum seimbang/ideal, dimana instrumen tes didominasi oleh butir soal dengan kategori mudah yaitu sebesar (50%).
4. Dilihat dari hasil analisis daya pembeda yang dilakukan, diketahui hanya terdapat 8 butir (16%) saja yang dinyatakan memiliki kualitas baik.

B. Saran

Berdasarkan pelaksanaan dan hasil penelitian ini, diajukan saran sebagai berikut.

1. Penyusun instrumen perlu memperhatikan dengan baik prosedur penyusunan tes pilihan ganda, agar tes pilihan ganda yang dibuat dapat dipertanggungjawabkan kualitasnya.
2. Soal yang memiliki kelemahan baik dari segi materi, konstruksi, dan bahasa yang digunakan sebaiknya diperbaiki agar kemampuan instrumen tes untuk melihat pencapaian kompetensi siswa menjadi lebih optimal.
3. Penelitian ini memungkinkan untuk dilanjutkan dengan lingkup penelitian yang lebih luas. Lingkup penelitian yang dimaksud dapat berupa lingkup sekolah, banyaknya peserta tes, dan metode analisis instrumen yang lain.

DAFTAR PUSTAKA

- Arifin, Zaenal. 2014. *Evaluasi Pembelajaran*. Bandung: Remaja Rosdakarya Offset.
- Arikunto, Suharsimi. 2009. *Dasar-dasar Evaluasi Pendidikan (Edisi Revisi)*. Jakarta: Rineka Cipta.
- _____. 2012. *Prosedur Penelitian: Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.
- Azwar, Saifuddin. 2007. *Tes Prestasi: Fungsi dan Pengembangan Pengukuran Prestasi Belajar*. Yogyakarta: Pustaka Pelajar
- Crocker, L. & Algina, J. 2005. *Introduction to Classical and Modern Test Theory*. Florida: Harcourt Brace Jovanovich College Publishers.
- Jihad, Asep dan Abdul Haris. 2013. *Penilaian Pembelajaran*, Yogyakarta: Multi Pressindo.
- Kusaeri dan Suprananto. 2012. *Pengukuran dan Penilaian Pendidikan*. Yogyakarta: GRAHA ILMU.
- Ngalim Purwanto, M. 2002. *Prinsip-prinsip dan Teknik Evaluasi Pengajaran*. Bandung: PT Remaja Rosdakarya.
- Nurgiyantoro, Burhan. 2011. *Penilaian Pembelajaran Bahasa*. Yogyakarta: BPF
- Purwanto. 2009. *Evaluasi Hasil Belajar*. Yogyakarta: Pustaka Pelajar
- Sudijono, Anas. 2006. *Pengantar Penilaian Pendidikan Ed. 1-6*. Jakarta: Raja Grafindo Persada.
- Sudjana, Nana. 2013. *Penilaian Hasil Proses Belajar Mengajar*, Bandung: Remaja Rosdakarya

- Sujarwadi, S. 2011. *Validitas dan Reliabilitas Instrumen Penelitian*. Jakarta: Universitas Negeri Jakarta.
- Sukardi. 2008. *Evaluasi Pendidikan Prinsip dan Operasional*. Jakarta: Bumi Aksara.
- _____. 2012. *Metodologi Penelitian Pendidikan: Kompetensi dan Praktiknya*. Yogyakarta: Bumi Aksara.
- Surapranata, S. 2004. *Analisis Validitas, Reliabilitas, dan Interpretasi Hasil Tes*. Bandung: Rosda.
- _____. 2009. *Analisis, Validitas, Reliabilitas, dan Intepretasi Hasil Tes: Implementasi Kurikulum 2004*. Bandung: PT Remaja Rosdakarya.
- Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 20 Tahun 2007 tentang Standar Penilaian Pendidikan. Jakarta.
- Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia Nomor 5 Tahun 2015 tentang Kriteria Kelulusan Peserta Didik, Penyelenggaraan Ujian Nasional, dan Penyelenggaraan Ujian Sekolah/Madrasah/Pendidikan Kesetaraan Pada SMP/MTS Atau yang Sederajat dan SMA/MA/SMK atau yang Sederajat. Jakarta.
- Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 23 Tahun 2006 tentang Standar Kompetensi Lulusan untuk Satuan Pendidikan Dasar dan Menengah. Jakarta.

Lampiran 12



KEMENTERIAN RISET, TEKNOLOGI DAN PENDIDIKAN TINGGI
UNIVERSITAS NEGERI SEMARANG (UNNES)
FAKULTAS ILMU SOSIAL (FIS)
 Gedung C.7 Kampus Sekaran, Gunungpati, Semarang 50229
 Website: fis.unnes.ac.id, E-mail: fis@unnes.ac.id, Telp./Fax. (024)8508006

Nomor : *3306* UN37.1.3/LT/2016
 Hal : Surat ijin Penelitian 01 JUN 2016

Yth. : Kepala Sekolah Madrasah Aliyah NU AI - Hikmah
 Kota Semarang

Dengan hormat, dalam rangka menyelesaikan proposal penelitian skripsi oleh mahasiswa di bawah ini:

Nama : Saekhul Indrianto
 N I M : 3201409053
 Semester : XIV (empatbelas)
 Jurusan/Prodi : Geografi / Pendidikan Geografi S1
 Jurusan/Fakultas : Geografi/ Ilmu Sosial
 Judul skripsi : Kualitas Instrumen Penilaian Ujian Madrasah Mata Pelajaran Geografi di Madrasah Aliyah NU AI – Hikmah Tahun Pelajaran 2015/2016
 Alokasi waktu : Bulan Mei s/d Agustus 2016

Mohon perkenan Saudara dapat mengizinkan mahasiswa dimaksud untuk melaksanakan Penelitian di Sekolah Madrasah NU AI - Hikmah , Kota Semarang

Atas kerjasamanya, disampaikan terima kasih.



an Dekan
Wakil Dekan Bidang Akademik
Prof. Dr. Wasino, M.Hum
 NIP. 19640805 1989011 001

Tembusan:
 1. Dekan
 2. Ketua Jurusan Geografi
 3. Kasubbag. Akd & Mawa
 4. Yang bersangkutan
 Fakultas Ilmu Sosial UNNES

FM-05-AKD-24/Rev.00

Lampiran 13


**LEMBAGA PENDIDIKAN MA'ARIF NU
MA NU AL HIKMAH SEMARANG**
 NSM / NIS : 131233740002 / 310160 NPSN : 20363044
 JL. KYAI AJI POLAMAN MIJEN KOTA SEMARANG 50217 JAWA TENGAH
 Phone : (024) 76671674 E-mail : manualhikmahsmg@gmail.com

SURAT KETERANGAN PENELITIAN
 NOMOR : 098/MA.NU.AH/VI/2016

Yang bertanda tangan di bawah ini Kepala MA NU AL – Hikmah, menerangkan bahwa:

Nama : Saekhul Indriyanto
 NIM : 3201409053
 Program Studi : Pendidikan Geografi S1
 Perguruan Tinggi : Universitas Negeri Semarang

Telah melakukan Observasi (Penelitian) di MA NU AL – Hikmah untuk keperluan Skripsi pada :

Waktu : Tanggal 1 Juni 2016
 Judul Skripsi : Kualitas Instrumen Penilaian Ujian Madrasah Mata Pelajaran Geografi di MA NU AL – Hikmah Tahun Pelajaran 2015/2016

Demikian Surat Keterangan ini dibuat untuk dapat dipergunakan sebagaimana mestinya.

Semarang, 30 Juni 2016
 Kepala Madrasah

 H. A. Futharno, S.Ag., M.Pd.I.


 UNIVERSITAS NEGERI SEMARANG



UNNES
UNIVERSITAS NEGERI SEMARANG