BAB II

LANDASAN TEORI, KERANGKA BERPIKIR DAN HIPOTESIS

2.1 Landasan Teori

2.1.1 Hakikat Evaluasi

Secara harfiah kata evaluasi berasal dari bahasa Inggris *evaluation*, dalam bahasa Indonesia berarti penilaian. Evaluasi bermanfaat dalam membuat pertimbangan dalam menilai sesuatu.

Di dalam evaluasi terdapat proses menggambarkan, memperoleh dan menyajikan informasi yang berguna untuk menilai keputusan. Sebagaimana diungkapkan oleh Stufflebeam dalam H.Daryanto, yaitu:

> Evaluation is the process of delineating, obtaining, and providing useful information for judging decision alternatives.[3]

Kegiatan pendidikan dan pembelajaran merupakan sebuah proses pencapaian sejumlah tujuan yang telah ditetapkan. Untuk mengetahui tercapainya tingkat keberhasilan tujuan pembelajaran diperlukan alat evaluasi. Tyler dalam Arikunto mengemukakan, bahwa:

---

[3] Stufflebeam dalam H. Daryanto, Evaluasi Pendidikan,(Jakarta: Rineka Cipta, 2007),h.1.

> Evaluasi sebagai sebuah proses pengumpulan data untuk menentukan sejauh mana,dalam hal apa,dan bagaimana tujuan pendidikan sudah tercapai.[4]

Pendidikan merupakan proses, evaluasi yang dilakukan juga harus merupakan proses, untuk mengukur kadar pencapaian tujuan yang telah ditentukan sebelumnya. Dengan evaluasi, guru mendapatkan informasi tentang hasil belajar siswa yang telah dicapai, sehingga guru dapat menentukan keberhasilan proses belajar mengajar yang diberikan.

Dari hasil evaluasi, siswa mengetahui sejauh mana kemampuannya dalam menguasai materi pembelajaran. Berdasarkan hal tersebut, kegiatan evaluasi mempunyai peran penting dalam pencapaian hasil belajar, untuk itu hendaknya evaluasi dipersiapkan dan dirancang dengan baik, sehingga proses belajar mengajar dapat berlangsung dengan baik dan mencapai hasil belajar yang baik pula.

2.1.2 Hakikat Tes

Untuk mengukur tujuan pendidikan perlu alat evaluasi. Sanjaya mengelompokan alat evaluasi menjadi dua teknik, yaitu tes dan nontes[5]. Alat evaluasi yang di-

---

[4] Ralph Tyler dalam Suharsimi Arikunto,op. cit.,(Jakarta: Bumi Aksara,2006)h.3.

gunakan biasanya berbentuk tes. Sudjana mendefinisikan tes sebagai berikut:

> Tes pada umumnya digunakan untuk menilai dan mengukur hasil belajar siswa, terutama hasil belajar kognitif berkenaan dengan penguasaan bahan pengajaran sesuai dengan tujuan pendidikan dan pengajaran.[6]

Berdasarkan hal tersebut, tes adalah penilaian yang digunakan untuk mengukur kemampuan, dan hasil belajar siswa dalam pencapaian suatu kompetensi tertentu sesuai dengan tujuan pendidikan. Sebuah tes dapat dikatakan benar-benar mengukur prestasi belajar siswa bila memberikan informasi tentang hasil belajar siswa sesuai dengan keadaan yang sebenarnya. Dengan kata lain tes diharapkan mampu memberikan informasi yang dapat dipertanggungjawabkan kebenarannya.

Tes yang digunakan sebagai alat ukur dalam pencapaian hasil belajar siswa adalah, tes yang dapat dipertanggungjawabkan, berkualitas dan sesuai dengan kriteria tertentu. Oleh karena itu, Lienert menyebutkan

---

[5] Wina Sanjaya, Pembelajaran dan Implementasi Kurikulum Berbasis Kompetensi,(Jakarta:Kencana, 2008), h. 187

[6] Nana Sudjana, Penilaian Hasil Proses Belajar Mengajar (Bandung: PT Remaja Rosdakarya, 2009), h.35

beberapa kriteria tes yang berkualitas baik ke dalam dua bagian,yaitu *Hauptgütekriterien*:

> Ein guter Test soll als *Hauptgütekriterien* folgende drei Forderungen erfüllen: er soll objektiv, reliable, und valid sein.[7]

Selanjutnya, *Nebengutekriterien*: ”er soll normiert, vergleichbar, ökonomisch, und nützlich sein.[8]”

Sebuah tes dikatakan baik sebagai alat ukur, bila memenuhi tiga kriteria *Hauptgütekriterien,* yaitu objektif, reliabel, valid. Serta empat kriteria *Nebengütekriterien*, yaitu baku, dapat dibandingkan, ekonomis dan bermanfaat.

Sesuai sudut pandang yang digunakan, tes yang biasa digunakan sebagai alat evaluasi ada beberapa macam. Diantaranya dilihat dari segi bentuk tes. Albers dan Bolton membagi bentuk tes berdasarkan cara peserta tes memberikan menjawab tes, yaitu: *offene Aufgaben* (tes terbuka), *halboffene Aufgaben* (tes setengah terbuka), dan

---

[7] Lienert und Raatz, Testaufbau und Testanalyse, (Weinheim: Psychologie verlags Union,1998) h. 7

[8] Ibid.

geschlossene Aufgaben (tes tertutup)[9], berikut definisi *offene Aufgaben*:

> Offen ist eine Aufgabe dann, wenn die Antwort relativ frei ausgeführt werden kann und von dem Kandidaten oder der Kandidatin als produktiv Leistung selbst formuliert werden muss….[10]

Dari definisi di atas, *offene Aufgaben* (tes terbuka) merupakan tes yang memberikan kebebasan pada peserta tes untuk menjawab, sehingga peserta tes secara produktif dapat memformulasikan jawaban sendiri. Contoh-nya pada tes uraian.

Sedangkan *halboffene Aufgaben* atau tes setengah terbuka:

> Halboffene Aufgabentyp ist die Antwort ebenfalls von den Kandidaten selbst zu formulieren, aller-dings innerhalb eines genau begrenzten Kontextes. z.B:Ergänzungsaugaben,Lückentest, Closetest[11]

Berdasarkan hal di atas, dalam *halboffene Aufgaben* peserta tes dapat memformulasikan jawaban sendiri. Akan tetapi, jawaban tersebut masih dibatasi konteks yang ada, contohnya pada tes bentuk isian.

---

[9] Hans-Georg Albers und Sibylle Bolton, Testen und Prüfen in der Grundstufe, (Berlin:Langenscheidt, 1999),h.28

[10] Ibid.

[11] Ibid.

Berbeda dengan *halboffene Aufgaben,* pada *geschlossene Aufgaben* atau tes tertutup, peserta tes memilih jawaban yang benar dengan memasangkan beberapa pilihan jawaban yang disediakan. Contohnya yaitu: pilihan ganda, tes benar-salah, tes menjodohkan, dan tes isian (tes rumpang). Hal ini seperti yang diungkapkan oleh Sibylle Bolton,

> …den geschlossenen Aufgaben, müssen die Schüler unter den vorgegeben Antworten die richtige Anwort auswählen. z.B. Multiple Choice Aufgaben, Ja oder Nein, Richtig oder Falsch Aufgaben und Zuordnungsaufgaben. [12]

Oleh karena itu, dalam melakukan evaluasi guru memerlukan alat ukur atau tes untuk menilai hasil belajar siswa. Untuk hasil tes yang baik, diperlukan tes yang baik, dan tes tersebut harus objektif, reliabel, valid. Selain itu, terdapat kriteria tambahan yaitu: standar, ekonomis, dapat dibandingkan, bermanfaat, dan dapat dipertanggungjawabkan. Terdapat beberapa bentuk tes: tes terbuka yaitu tes uraian dan tes tertutup yaitu tes pilihan ganda, tes menjodohkan, dan tes benar-salah.

Dalam penelitian ini, dipilih bentuk tes tertutup yaitu tes pilihan ganda, karena tes pilihan ganda adalah

---

[12] Ibid., h. 30

tes yang selalu digunakan guru pada ulangan blok di sekolah.

2.1.2.1 Tes Pilihan Ganda

Tes bentuk tertutup atau tes obyektif memiliki model yang lebih banyak dan variatif dibandingkan tes bentuk uraian, karena itulah tes tertutup lebih sering digunakan guru dalam mengukur hasil belajar siswa dalam menguasai materi pelajaran. Bolton menyatakan bahwa:

> Bei Multiple-choice-Aufgaben(häufig auch Mehr-fachwahl-Aufgaben genannt)müssen die Lernenden aus mehreren vorgegebenen Antworten die richtige Antwort auswählen.[13]

Tes pilihan ganda merupakan bentuk tes tertutup. Siswa diminta untuk memilih satu jawaban yang benar dari beberapa pilihan jawaban yang telah disediakan.

Dalam pembuatan tes pilihan ganda terdapat bebe-rapa struktur yang harus diperhatikan, yaitu: memiliki sebuah pernyataan, terdapat pilihan jawaban (alternatif), jawaban benar dan distraktor. Di dalam sebuah tes pilihan ganda setidaknya disediakan lebih dari satu pilihan

---

[13] Sybille Bolton, Problem der Leistungsmessung Lernfortschrittsests in der Grundstufe, (München: Langenscheidt, 1996), h. 31

jawaban. Hal ini sejalan dengan yang diungkapkan oleh Sibyle Bolton, yaitu:

> Als mögliche Antworten auf die Frage oder Forsetzung des einleitenden Satzes werden mindestens zwei, höchstens in der Regel vier Alternativen angeboten. Nur eine der Alternativen ist im allgemeinen richtig(die Antwort), alle anderen sind falsch(distraktoren)[14].

Atas sebuah pertanyaan tes pilihan ganda setidaknya minimal disediakan dua pilihan jawaban dan maksimal empat pilihan jawaban. Hanya ada satu jawaban benar dari pilihan jawaban yang ada, dan semua jawaban yang salah berfungsi sebagai pengecoh(distaktor).

Dalam mengolah skor dalam tes bentuk pilihan ganda, diperlukan dua macam rumus, seperti diungkapkan Arikunto, yaitu: dengan denda dan tanpa denda[15]. Cara penskoran dengan denda adalah dengan hukuman karena diragukan adanya unsur tebakan. Rumus penskoran dengan menggunakan denda[16]:

$$S = R - \frac{W}{0-1}$$

S = Skor yang diperoleh (*Raw Score)*

---

[14] Sybille Bolton, op.cit., h.30

[15] Arikunto, op. cit., h.172

[16] Ibid.

R = Jawaban yang betul

W = Jawaban yang salah

0 = banyaknya *option/*pilihan

1 = bilangan tetap

Atau dapat menggunakan Cara penskoran tanpa denda adalah apabila banyaknya angka yang diperoleh siswa sebanyak jawaban yang cocok dengan kunci. Rumus penskoran tanpa denda[17]: S = R

Bentuk tes pilihan ganda ini sering dipergunakan oleh para guru dibanding bentuk tes lain sebagai alat evaluasi. Hal ini dikarenakan bentuk tes pilihan ganda memepunyai beberapa kelebihan dibanding bentuk tes lain. Hal ini seperti yang dipaparkan oleh Arikunto, antara lain:

> Kelebihan tes pilihan ganda adalah lebih representatif mewakili isi dan luas bahan, lebih mudah dan cepat cara mengkoreksinya, pemeriksaannya dapat diserahkan orang lain dan dalam pemeriksaan tidak ada unsur subjektif yang mempengaruhi[18].

Tes pilihan ganda merupakan salah satu bentuk tertutup, yang terdiri dari pertanyaan dan beberapa pilihan jawaban sebagai pengecoh dan satu jawaban benar.

---

[17] Ibid.

[18] Ibid.,h.164

Tes pilihan ganda memiliki beberapa kelebihan, yaitu lebih representatif mewakili isi dan luas bahan, lebih mudah dan cepat mengoreksinya, pemeriksaannya dapat diserahkan orang lain dan tidak subjektif.

Dalam penelitian ini, penulis memaparkan teori mengenai tes pilihan ganda, karena objek diteliti adalah hasil jawaban dari tes bentuk pilihan ganda. Dalam penelitian ini tidak menggunakan kedua rumus penskoran, karena hasil skor telah diketahui. Jenis tes akan diteliti adalah bentuk tes pilihan ganda pada Ulangan Blok semester 1 Bahasa Jerman di SMA Jakarta.

2.1.3 Hakikat Analisis Butir Soal

Untuk mengetahui apakah tes yang diberikan, atau yang telah dibuat sudah sesuai prosedur sistematis dan berkualitas, diperlukan sebuah analisis butir soal. Hal ini seperti yang diungkapkan oleh Sudjana:

> Analisis butir soal atau analisis item adalah pengkajian pertanyaan-pertanyaan tes agar diperoleh perangkat pertanyaan yang memiliki kualitas yang memadai.[19]

Berdasarkan hal tersebut, dalam analisis butir soal terdapat identifikasi terhadap tes yang diberikan untuk

---

[19] Nana Sudjana, op.cit, h. 135

mendapatkan sebuah perangkat tes agar mempunyai kualitas yang layak. Analisis butir soal mempunyai tujuan antara lain untuk mengadakan identifikasi soal-soal yang baik, kurang baik, dan soal yang jelek[20]. Analisis butir soal merupakan pengkajian tes untuk mengidentifikasi soal, agar diperoleh perangkat tes yang memiliki kualitas soal yang memadai.

Analisis butir soal dalam penelitian ini dilakukan dengan cara kuantitatif atau validitas empiris (*empirical validity*). Sumarna menjelaskan bahwa dalam analisis butir soal kuantitatif cara menganalisisnya lebih ditekankan pada karakteristik isi tes tersebut melalui data yang diperoleh secara empiris.[21] Dalam penelitian ini dilakukan analisis butir soal kuantitatif dengan analisis karakter isi tes, karena analisis butir soal kuantitatif dapat dilakukan untuk bentuk tes pilihan ganda.

Arikunto menjelaskan, bahwa dalam analisis butir soal untuk soal pilihan ganda, analisis soal menggunakan tingkat kesulitan soal, daya pembeda,pola jawaban soal,

---

[20] Suharsimi Arikunto, op. cit, h.207

[21] Sumarna Supranata, Analisis Validitas, Reliabilitas, dan Interpretasi Hasil Tes, (bandung: PT. Remaja Rosdakarya, 2006), h. 1

selain validitas dan reliabilitas.[22] Berdasarkan hal tersebut, analisis butir soal adalah kegiatan pengkajian soal dalam bentuk pilihan ganda yang dilakukan dengan membuat analisis hitung tingkat kesulitan, daya pembeda, dan pola jawaban soal, selain validitas dan reliabilitas, untuk mendapatkan kriteria soal yang baik, kurang baik, dan jelek.

2.1.3.1 Tingkat kesulitan

Analisis tingkat kesulitan adalah analisis untuk menentukan kriteria soal: termasuk mudah, sedang atau sukar dengan menggunakan rumus. Dalam menganalisis tingkat kesulitan soal terdapat beberapa kriteria. Arikunto membagi kriteria tersebut dengan kesulitan soal sebagai berikut:

> Soal dengan P 0,00 sampai 0,30 adalah soal sukar
> Soal dengan P 0,30 sampai 0,70 adalah soal sedang
> Soal dengan P 0,70 sampai 1,00 adalah soal mudah[23]

Soal dikatakan baik, apabila soal termasuk mempunyai tingkat kesulitan sedang, yaitu pada kisaran 0,30-0,70. Hal tersebut berbeda dengan yang diungkapkan

---

[22] Arikunto, op.cit., h.207

[23] Ibid., h.210

Sudjana. Menurut Sudjana, soal akan mempunyai kualitas baik jika soal tersebut mempunyai proporsi yang tepat antara soal kategori mudah, sedang, dan sulit.[24]

Ada beberapa perbandingan dalam menentukan proporsi jumlah soal kategori soal mudah, sedang, dan sulit. Menurut Sudjana perbandingan soal mudah-sedang-sulit dapat dibuat 3-4-3 (30% soal mudah, 40% soal sedang, 40% soal sulit) dan 3-5-2 (30% soal mudah, 50% soal sedang, 20% soal sulit).[25]

Berdasarkan hal tersebut, soal yang baik adalah soal yang mempunyai tingkat kesulitan yang tidak terlalu mudah dan tidak terlalu sulit, atau soal yang termasuk kategori sedang, dan memiliki proporsi tingkat kesulitan soal yang tepat. Hal ini karena soal yang mudah tidak akan merangsang siswa untuk memecahkannya, sedangkan soal yang terlalu sulit akan menyebabkan siswa menjadi putus asa untuk mencobanya.

---

[24] Nana Sudjana, op. cit., h.135

[25] Ibid., h. 135-136

2.1.3.2 Daya Pembeda

Daya pembeda soal adalah mengkaji butir-butir soal untuk mengetahui kesanggupan soal membedakan siswa yang tergolong pandai dan kurang pandai[26]. Artinya dari pengelompokan berdasarkan skor siswa yang diperoleh, dapat diketahui siswa yang termasuk kelompok *Upperclass* (Kelompok tinggi) dan kelompok *Lowerclass*(kelompok rendah).

Bilamana siswa kelompok tinggi menjawab benar suatu butir soal dan kelompok rendah menjawab salah, hal ini berarti butir soal yang bersangkutan memiliki daya pembeda yang baik. Sebaliknya jika terjadi kelompok rendah menjawab benar lebih banyak daripada kelompok tinggi berarti butir soal tersebut mempunyai daya pembeda yang tidak baik.

Hal ini menandakan butir soal tersebut kurang dapat membedakan bahkan tidak dapat membedakan siswa yang pandai dan yang kurang pandai, untuk itu butir soal tersebut perlu direvisi atau diganti. Untuk mengetahui dan membedakan kemampuan antara siswa yang pandai, sedang

---

[26] Ibid., h. 141

dan yang kurang, disusunlah sebuah klasifikasi daya pembeda. Arikunto mengklasifikasikannya sebagai berikut:

D: 0,00-0,20 = jelek
D: 0,20-0,40 = cukup
D: 0,40-0,70 = baik
D: 0,70-1,00 = baik sekali
D: negatif = semuanya tidak baik[27]

Soal dengan klasifikasi negatif merupakan soal yang tidak baik. Butir soal yang dikatakan baik dalam daya pembeda menurut Arikunto adalah butir soal yang berada pada 0,40 sampai dengan 0,70[28]. Sebuah soal di-katakan memiliki daya pembeda baik jika soal tersebut memiliki lebih dari 0,40 sampai dengan 1,00, akan tetapi bila soal tersebut memiliki daya pembeda kurang dari 0,40 maka soal tersebut dikatakan jelek.

2.1.3.3 Pola Jawaban Soal

Pola jawaban soal adalah distribusi peserta tes *(testee)* dalam hal menentukan pilihan jawaban pada soal bentuk pilihan ganda. Menurut Daryanto pola jawaban soal diperoleh dengan menghitung banyaknya peserta tes(*testee*)

---

[27] Suharsimi Arikunto, op. cit., h. 218

[28] Ibid.

yang memilih pilihan jawaban a,b,c,d atau yang tidak memilih pilihan manapun yang disebut dengan omit (O)[29].

Dari pola jawaban dapat ditentukan pengecoh (*distractor*) berfungsi dengan baik, atau tidak baik. Pengecoh yang tidak dipilih sama sekali oleh peserta tes berarti pengecoh itu jelek. Dengan melihat pola jawaban soal dapat diketahui:

- Taraf kesukaran
- daya pembeda soal
- baik dan tidaknya distraktor[30]

Menurut Nitko, ada beberapa kriteria dalam soal pilihan ganda, yaitu dilihat dari nilai koefisien tingkat kesulitan, daya pembeda dan proporsi jawaban atau yang biasa disebut pola jawaban soal.[31] Berikut tabel kriteria pemilihan soal pilihan ganda menurut aturan Nitko dalam Supranata.

---

[29] H. Daryanto, Evaluasi Pendidikan, (Jakarta, Rineka Cipta, 2007), h.192

[30] Suharsimi Arikunto, op.cit., h.220

[31] Nitko dalam Sumarna Supranata, op. cit. h.47

Kriteria Pemilihan soal Pilihan Ganda[32]

Kriteria	Koefisien	Keputusan
Tingkat Kesulitan	0,30 s.d 0,70 0,10 s.d 0,29 atau 0,70 s.d 0,90 < 0,10 dan > 0,90	Diterima Direvisi Ditolak
Daya Pembeda	> 0,30 0,10 s.d 0,29 < 0,10	Diterima Direvisi Ditolak
Proporsi jawaban	> 0,50	

[32] Ibid.

Sumber : Nitko dalam Sumarna Supranata, Analisis Validitas, Reliabilitas, dan Interpretasi Hasil Tes, (Bandung :PT. Remaja Rosdakarya, 2005), h.47

Berdasarkan tabel diatas kriteria soal yang di-terima salah satunya memiliki proporsi jawaban memiliki nilai lebih dari 0,50. Hal ini sejalan dengan Arikunto, sebuah distraktor dikatakan baik jika paling sedikit dipilih 5% seluruh peserta tes.[33]

Dalam menganalisis butir soal selain mencari tingkat kesulitan, daya pembeda soal dan pola pilihan jawaban, juga mencari validitas dan reliabilitas butir soal merupakan termasuk dalam menganalisis butir soal.

2.1.3.4 Validitas

Agar diperoleh data yang baik, maka instrument atau tes harus valid. Untuk mengetahui hal tersebut di-perlukan analisis validitas. Analisis validitas bertujuan untuk mengkaji kesahihan alat ukur atau soal dalam me-nilai apa yang seharusnya diukur atau mengkaji ketepatan soal tes sebagai alat ukur.

Dalam mengukur validitas kesahihan atau ketepatan suatu tes sebagai alat ukur adalah yang terpenting. Dengan kata lain, tes harus benar-benar mengukur apa yang

---

[33] Suharsimi Arikunto, op.cit., h. 220

hendak diukur. Tes dirancang untuk mengukur seberapa jauh siswa mencapai suatu tujuan pembelajaran siswa.

Untuk menentukan apakah sebuah butir soal valid, maka setelah diperoleh koefisien, lalu dibandingkan dengan tabel *Product moment*. Seperti yang dijelaskan oleh Arikunto bahawa untuk mencari validitas adalah dengan teknik korelasi *product moment* dengan simpangan atau dengan angka kasar yang dikemukakan oleh Pearson.[34]

Koefisien korelasi umumnya dibagi ke dalam lima bagian, berikut interpretasi mengenai besarnya koefisien korelasi:

> Antara 0,800 sampai dengan 0,1000= sangat tinggi
> Antara 0,600 sampai dengan 0,800= tinggi
> Antara 0,400 sampai dengan 0,600= cukup
> Antara 0,200 sampai dengan 0,400= rendah
> Antara 0,00 sampai dengan 0,200= sangat rendah[35]

Dari indeks di atas dapat diketahui jika besar koefisien korelasi antara 0,400 sampai dengan 0,1000 dapat dikata-kan baik, tetapi jika besar koefisien korelasi kurang dari 0,400 dikatakan tidak baik. Jika harga r kurang dari

---

[34] Suharsimi Arikunto, op. cit, h.69
[35] Arikunto, op. cit., h. 75

harga *product moment* dalam tabel, korelasi tersebut tidak signifikan. Begitu juga sebaliknya.[36]

Dalam menafsirkan harga korelasi terdapat dua cara:

> Dengan melihat harga r dan diinterpretasikan. Kemudian dengan berkonsultasi ke tabel harga kritik *r product moment* sehingga diketahui signifikan tidaknya korelasi tersebut.[37]

Dalam penelitian ini perhitungan validitas data menggunakan rumus korelasi *product moment* dengan rumus angka kasar(dapat dilihat pada halaman 36), karena dengan rumus angka kasar bilangannya besar-besar dan bulat(contoh pada halaman 65). Harga koefisien korelasi yang didapat kemudian dikonsultasikan pada tabel harga kritik *product moment*.

2.1.3.5 Reliabilitas

Reliabilitas merupakan hal pokok dalam sebuah penelitian yang tidak boleh ditinggalkan. Kerlinger mendefinisikan reliabilitas, sebagai berikut:

> Reliabilität wird auch als Zuverlässigkeit, Genauigkeit und Stabilität der Messung bezeichnet. Sie liegt vor, wenn wiederholte

---

[36] Ibid.

[37] Ibid.

> Messvorgänge unter sonst gleichen Rahmen-bedingungen zu gleichen Werten führen.[38].

Dari definisi di atas, reliabiltas juga dikenal sebagai keandalan, akurasi, dan stabilitas dari peng-ukuran. Sesuatu dapat dikatakan reliabel bila pengukuran ulang yang dilakukan dengan kondisi yang sama akan meng-hasilkan nilai yang sama.

Dalam penelitian ini reliabilitas tes digunakan guru sebagai alat evaluasi. Sebuah tes dikatakan dapat dipercaya, jika tes tersebut mempunyai hasil yang tetap (ajeg). Menurut Anderson dalam Arikunto:

> *a reliable measure in one that provides consistent and stable indication of the characteristic being investigated*[39]*.*

Dapat dikatakan bahwa sebuah pengukuran yang tetap (ajeg), tentu memberikan indikasi karakteristik yang konsisten dan stabil. Oleh karena itu untuk mendapat reliabilitas tes yang baik, tes tersebut harus diujikan kepada siswa yang sama lebih dari satu kali, atau dari dua perangkat tes yang setara kepada objek yang sama.

---

[38] Kerlinger, http://www.wirtschaftslexikon24.net/d/reliabilitaet/reliabilitaet.htm, diakses tanggal 27 November 2009, pukul 19.05 WIB.

[39] Scarvia B. Anderson dalam Suharsimi Arikunto, op. cit., h. 87

Arikunto menyebutkan ada tiga cara untuk mencari reliabilitas, yaitu metode bentuk pararel (*equivalent*), metode tes ulang (*test-retest method*), dan metode belah dua (*splithalf method*)[40]. Untuk mencari taraf reliabilitas tanpa harus melakukan tes dua kali, digunakanlah metode belah dua (*splithalf method*). Arikunto menjelaskan bahwa,

> Dalam menggunakan metode ini pengetes hanya menggunakan sebuah tes dan dicobakan satu kali. Pada waktu membelah dua dan mengkorelasikan dua belahan, baru diketahui reliabilitas separo tes. Untuk mengetahui reliabilitas keseluruhan tes harus menggunakan rumus Spearman Brown.[41]

Butir soal dikatakan reliabel, jika nilai r hitung lebih tinggi atau sama dengan nilai r tabel, sedangkan butir soal yang tidak reliabel adalah butir soal yang nilai r hitung kurang dari nilai r tabel.

Dalam penelitian ini, metode yang dipakai untuk mencari reliabilitas adalah dengan menggunakan metode belah dua. Metode ini dipakai karena tes dilaksanakan hanya satu kali. Kemudian hasil tes dianalisis reliabilitasnya dengan membelah dua hasil tes tersebut dan dikorelasikan dengan menggunakan persamaan, hasil hitung reliabilitas yang didapat akan dikonsultasikan dengan

---

[40] Ibid., h.90-92

[41] Ibid., h.92-93

tabel harga kritik *r product moment.*

2.1.3.6 Tabel Analisis Butir Soal

Menurut penjelasan kriteria analisis butir soal diatas berdasarkan tingkat kesulitan soal, daya pembeda soal, pola jawaban soal, validitas soal dan reliabilitas soal. Penulis membuat tabel analisis butir soal berdasarkan teori-teori yang telah disebutkan di atas. Tabel analisis butir soal tersebut sebagai berikut:

Tabel 1. Analisis Butir Soal

Tingkat Kesulitan			Daya Pembeda		Pola Jawaban		Validitas		Reliabilitas	
IK≤0,30	0,31≤IK≥0,70	IK≥0,71	ID<0,40	ID>0,40	PJ<5%	PJ>5%	V<t	V>t	r<t	r>t

2.2 Hasil Penelitian yang Relevan

Peneliti telah menemukan beberapa hasil penelitian yang serupa tentang analisis butir soal diantaranya yaitu:

1. Muthia Tri Handayani, Analisis Kualitas Butir-Butir Soal Tes hasil Belajar Matematika Kelas III CAWU II di SLTP 110 Jakarta, Jakarta: Jurusan Teknologi Pendidikan, FIP, UNJ, 2000[42]
2. Liya Wahyuningsih, Analisis Kesesuaian Butir Soal Tes Sumatif Buatan Guru Mata Pelajaran Bahasa Indonesia SMP Kelas VII berdasarkan KBK,Jakarta: Jurusan Bahasa dan Sastra Indonesia, FBS, UNJ, 2006[43]

---

[42] Muthia Tri Handayani, Analisis Kualitas Butir-Butir Soal Tes hasil Belajar Matematika Kelas III CAWU II di SLTP 110 Jakarta, Jakarta: Jurusan Teknologi Pendi-dikan, FIP, UNJ, 2000

[43] Liya Wahyuningsih, Analisis Kesuaian Butir Soal Tes Sumatif Buatan Guru Mata Pelajaran Bahasa Indonesia SMP Kelas VII berdasarkan KBK,Jakarta: Jurusan Bahasa dan Sastra Indonesia, FBS, UNJ, 2006

3. Kartini Megawati, Kualitas Soal Ulangan Umum Semester I Bahasa Jerman Kelas X SMAN 48 Jakarta Tahun Pelajaran 2005/2006, Jakarta: Jurusan Pendidikan Bahasa Jerman, FBS, UNJ, 2007[44]

2.3 Kerangka Berpikir

Evaluasi merupakan proses dalam memberikan pertimbangan dalam mengukur apakah tujuan pendidikan sudah tercapai. Untuk itu diperlukan alat ukur yaitu tes. Tes yang digunakan sebagai alat ukur harus tes yang berkualitas sehingga tes dapat memberikan informasi hasil belajar.

Selayaknya tes digunakan untuk mengukur hasil belajar, akan tetapi hasil tes kadang tidak menunjukan hasil yang ditargetkan guru. Hasil tes yang jelek belum tentu dikarenakan kemampuan siswa yang rendah, hal ini mungkin dikarenakan butir soal dalam tes tidak mengukur kemampuan hasil belajar yang telah diperoleh siswa. Oleh karena itu, diperlukan analisis butir soal untuk mengetahui kualitas soal pada tes tersebut.

---

[44] Kartini Megawati, Kualitas Soal Ulangan Umum Semester I Bahasa Jerman Kelas X SMAN 48 Jakarta Tahun Pelajaran 2005/2006, Jakarta: Jurusan Pendidikan Bahasa Jerman, FBS, UNJ, 2007

Tes yang mempunyai kualitas baik dan mempunyai kriteria yang baik akan menghasilkan sebuah tes tes yang baik dan dapat dipertanggungjawabkan. Oleh karena itu, diperlukan sebuah analisis butir soal untuk mengadakan identifikasi soal-soal yang baik, kurang baik, dan soal yang jelek.

Dalam penelitian ini, penulis menggunakan analisis butir soal kuantitatif dengan menganalisis tingkat kesulitan, daya pembeda, pola pilihan jawaban, validitas dan reliabilitas. Pada penelitian ini penulis menganalisis butir soal ulangan blok semester 1 Bahasa Jerman kelas XII SMA Negeri Jakarta. Data penelitian adalah 350 lembar jawaban komputer siswa pada ulangan blok semester 1 Bahasa Jerman kelas XII SMA Negeri Jakarta.

Setelah data dianalisis dengan mencari tingkat kesulitan, daya pembeda dan pola pilihan jawaban, dicari validitas dan reliabilitas masing-masing butir soal. Menghitung validitas butir soal digunakan teknik korelasi product moment Pearson dengan angka kasar. Setelah diketahui validitas, kemudian dihitung reliabilitas butir soal dengan menggunakan metode belah dua, dengan membagi dua data awal dan akhir, dan dihitung menggunakan rumus Spearman Brown, hasil hitung yang didapat akan dibanding-

kan dengan tabel *product moment.*

Hasil hitung akan dimasukkan ke dalam tabel analisis. Selanjutnya hasil penelitian diinterpretasikan dan ditarik kesimpulan. Dari penelitian ini akan diketahui berapa butir soal yang berkualitas dalam Ulangan Blok semester I Bahasa Jerman kelas XII di SMA Negeri Jakarta.

2.4 Hipotesis Penelitian

Dari kerangka berpikir yang telah dikemukakan diatas dapat dirumuskan yaitu:

Ha = Butir soal ulangan blok semester I Bahasa Jerman kelas XII di SMA berkualitas baik.