

BAB II

TINJAUAN PUSTAKA

A. Deskripsi Teoretis

A.1 Evaluasi

Sebelum membicarakan ujian akhir semester perlu dibicarakan mengenai evaluasi karena ujian akhir semester merupakan bentuk dari evaluasi. Kata evaluasi berasal dari bahasa Inggris *evaluation*. Dalam Kamus Besar Bahasa Indonesia kata evaluasi berarti penilaian. Menurut Stockmann (2004:2)

“Evaluationen stellen ein wichtiges Instrument zur Generierung von Erfahrungswissen dar. Sie werden durchgeführt in dem Informationen gesammelt und anschließend bewertet werden, um letztendlich Entscheidungen zu treffen.”

Disebutkan bahwa evaluasi merupakan instrumen penting untuk mengolah data dan digunakan untuk mengumpulkan informasi dan menilainya, sehingga dapat diambil keputusan berdasarkan informasi tersebut. Selain itu Stamm (2003:140) mengemukakan hal-hal ada pada proses evaluasi yaitu, *“Zum Evaluationsprozess gehört die Wahl eines angemessen Designs, die Qualität der Datenerhebung, die Präsentation der Daten, sowie die Kongruenz zwischen die Zielsetzungen und Resultaten.”* Stamm menyebutkan bahwa dalam proses evaluasi terdapat pemilihan desain evaluasi yang tepat (alat ukur evaluasi), kualitas pengumpulan data (kualitas alat ukur), penyajian data, serta kesesuaian antara tujuan dan hasil evaluasi tersebut. Sedangkan dalam kaitan pada bidang pendidikan, Sukardi

(2009:2) menyatakan bahwa evaluasi merupakan cara untuk menilai kemampuan siswa dalam proses belajar mengajar. Artinya evaluasi umumnya diberikan oleh guru kepada siswa agar kemajuan siswa dalam proses pembelajaran dapat diketahui. Senada dengan Sukardi, Tyler dalam Arikunto (2006:3) juga menyatakan bahwa proses evaluasi dilakukan untuk menentukan sejauh mana, dalam hal apa, dan bagaimana tujuan pendidikan sudah tercapai. Scholz (2013:74) juga menyatakan, *“Anhand der Evaluationsergebnisse wird überprüft, in wieweit das Verhältnis von angestrebten Zielen und den tatsächlich erzielten Erfolgen übereinstimmt”*. Disebutkan bahwa berdasarkan hasil evaluasi dapat diketahui sejauh mana hubungan dari tujuan dan keberhasilan proses belajar sudah tercapai. Artinya dalam dunia pendidikan, proses evaluasi merupakan hal yang wajib dilakukan untuk mengetahui apakah tujuan pendidikan yang dilakukan oleh peserta didik sudah tercapai atau belum.

Evaluasi dalam proses belajar mengajar tidak hanya dilakukan untuk mengetahui pencapaian belajar siswa, namun juga untuk menilai sejauh mana sebuah proses belajar mengajar telah berjalan dengan baik. Hal ini sesuai dengan pernyataan Purwanto (2010:5) bahwa guru dapat melakukan kegiatan evaluasi untuk mengukur atau menilai keefektifan pengalaman, kegiatan, dan metode mengajar yang digunakan. Jadi yang dimaksud dengan evaluasi adalah proses mencari tahu hasil belajar siswa yang dapat digunakan oleh guru maupun siswa untuk melihat pencapaian siswa, sejauh mana penguasaan siswa pada materi yang telah diajarkan guru, juga kelemahan-kelemahan siswa dalam menguasai suatu materi serta untuk menilai keefektifan kegiatan belajar itu sendiri.

A.2 Tes

Tes merupakan salah satu bentuk dari evaluasi yang fungsinya adalah sebagai alat ukur. Hal ini sesuai dengan pernyataan Sudjana (2010:35) bahwa tes digunakan untuk menilai dan mengukur hasil belajar siswa melihat penguasaan siswa pada materi yang sesuai tujuan pendidikan dan pengajaran. Tes yang diberikan kepada siswa juga harus mempunyai kriteria yang baik. Arikunto (2006:58) menyatakan bahwa tes sebagai alat ukur yang baik harus memenuhi persyaratan yaitu: validitas atau ketepatan tes dalam mengukur apa yang harus diukur, reliabilitas atau ketetapan atau keajegan tes dalam menilai apa yang dinilainya, obyektivitas yang berarti tidak adanya faktor subyektif yang mempengaruhi tes dan sistem penilaiannya, praktikabilitas atau kemudahan pelaksanaan dan pemeriksaan tes, serta ekonomis yang berarti pelaksanaan tes yang tidak memerlukan biaya mahal. Hampir senada dengan Arikunto, Liener (1998:207) mengemukakan. *“Ein guter Test soll als Hauptgütekriterien folgende drei Forderungen erfüllen: “er soll objektiv, er soll reliable, er soll valide sein”*. Disebutkan bahwa tes yang baik juga harus memenuhi kriteria sebagai berikut: obyektif, reliabel dan valid. Oleh karena itu dapat disimpulkan bahwa tes merupakan alat evaluasi yang dipakai guru untuk menilai dan mengukur hasil belajar siswa yang bentuknya harus memenuhi syarat yaitu obyektif, reliabel dan valid.

A.2.1 Jenis Tes

Sudjana (2009:35) membagi jenis tes yang digunakan sebagai alat penilaian hasil belajar menjadi dua, yaitu tes uraian (esai) dan tes objektif.

1. Tes Uraian

Tes uraian atau tes *essay* adalah tes yang pertanyaannya menuntut siswa untuk menguraikan atau menjelaskan jawabannya sesuai dengan kemampuan dan bahasa siswa sendiri.

2. Tes Obyektif

Tes obyektif adalah tes dengan pertanyaan dengan pilihan jawaban yang sudah disediakan, sehingga siswa hanya perlu memilih jawaban yang paling tepat dari pilihan jawaban tersebut . Beberapa bentuk soal dari tes obyektif adalah bentuk pilihan benar-salah, pilihan ganda, menjodohkan, dan isian pendek atau melengkapi.

A.2.1.1 Tes Pilihan Ganda

Dalam penelitian ini, Ujian Akhir Semester bahasa Jerman kelas XII tahun ajaran 2012/2013 semester 1 di SMA Negeri 54 menggunakan bentuk soal pilihan ganda. Seperti yang telah dijelaskan sebelumnya bahwa salah satu bentuk tes obyektif adalah soal pilihan ganda. Soal pilihan ganda disebut juga *Multiple Choice* dalam bahasa Inggris atau *Mehrfach-Wahl-Aufgabe* dalam bahasa Jerman. Menurut Gronlund dalam Lal (2008:298), soal pilihan ganda terdiri dari masalah dan solusi-solusi yang ditawarkan. Masalah bisa berupa pertanyaan langsung atau

kalimat rumpang, sedangkan solusi yang ditawarkan dapat berupa kata, angka, simbol atau frasa yang disebut juga alternatif.

A multiple choice item consists of a problem and a list of suggested solutions. The problem may be stated as a direct question or an incomplete statement and is called the stem of the item. The text of the suggested solutions may include words, numbers, symbols or phrases and are called alternatives (also called choices or options).

Soal bentuk pilihan ganda mempunyai struktur khusus. Sudjana (2009:48) mengatakan mengatakan bahwa soal pilihan ganda terdiri dari *stem*, *option*, kunci dan distraktor.

Dilihat dari stukturanya, bentuk soal pilihan ganda terdiri atas: *Stem*: pertanyaan atau pertanyaan yang berisi permasalahan yang akan ditanyakan, *Option*: sejumlah pilihan atau alternatif jawaban, *Kunci*: jawaban yang benar atau paling tepat dan *Distractor* (pengecoh): jawaban-jawaban lain selain kunci jawaban.

Glaboniat (1998:102) menjelaskan keuntungan dari soal pilihan ganda yaitu, “*Die Vorteile des Multiple Choice Format liegen ebenfalls in der einfachen Auswertbarkeit und der (vermeintlichen) Objektivität des Verfahrens*”. Disebutkan bahwa keuntungan dari soal pilihan ganda terletak pada penilaiannya yang mudah dan obyektivitas pemeriksaannya. Hampir senada dengan Glaboniat, Sudjana (2009:49) menyebutkan kebaikan bentuk soal pilihan ganda adalah dapat mencakup materi yang banyak, lebih cepat dan mudah untuk dikoreksi karena menggunakan kunci jawaban, dan penilaiannya bersifat obyektif.

Dari penjelasan-penjelasan tersebut dapat disimpulkan bahwa soal pilihan ganda adalah salah satu bentuk tes obyektif yang dapat digunakan untuk evaluasi dengan materi-materi yang banyak dan karena bentuknya yang obyektif, guru dapat lebih mudah melakukan pemeriksaan atau pengkoreksian.

A.3 Analisis Butir Soal

Sudjana (2009:135) menyatakan bahwa analisis butir soal adalah pengkajian butir soal dalam tes agar diperoleh soal dengan kualitas yang memadai. Sementara itu menurut Arikunto (2006:207) mengatakan bahwa tujuan analisis butir soal adalah untuk mengidentifikasi soal-soal yang baik, kurang baik dan jelek. Senada dengan Sudjana dan Arikunto, menurut Bortzt dan Döring (2006:730), *“Itemanalyse (Aufgabenanalyse) Überprüfung der Gütekriterien einzelner Items, um die Qualität eines Tests oder Fragebogens einschätzen zu können und ggf. Durch Austausch oder Veränderung einzelner Items (Testrevision) zu Verbesserungen zu kommen.”* Disebutkan bahwa analisis butir soal adalah pengkajian kriteria-kriteria kualitas sebuah item, untuk menilai kualitas sebuah tes atau kuesioner dan jika diperlukan dapat dilakukan dengan mengubah atau menukar masing-masing item (revisi) untuk memperbaiki item tersebut. Oleh karena itu dapat disimpulkan bahwa analisis butir soal merupakan kegiatan pengkajian suatu tes yang bertujuan untuk mengetahui kualitas masing-masing butir soal dalam tes tersebut.

Secara umum hal-hal yang harus dicari atau dianalisis dalam proses analisis butir soal adalah tingkat validitas, reabilitas, kesukaran, dan daya pembeda serta fungsi pengecoh.

A.3.1 Validitas

Arikunto (2006:59) menyatakan bahwa tes disebut valid apabila tes itu dapat mengukur apa yang hendak diukur. Senada dengan Arikunto, Sudjana

(2010:149) menyatakan bahwa tujuan analisis validitas adalah mengkaji kesahihan alat ukur atau tes untuk menilai apa yang seharusnya diukur atau mengkaji ketepatan soal tes yang diukur. Hal yang sama diungkapkan oleh Bühner (tt:36) yang mengatakan bahwa pengertian validitas adalah sejauh mana sebuah tes mengukur apa yang diukur, *“unter Validität versteht man im eigentlichen Sinne das Ausmaß, in dem ein Test das misst, was es zu messen vorgibt”*.

Arikunto (2006:65) menyatakan bahwa secara garis besar ada dua macam validitas, yaitu validitas logis dan validitas empiris. Penjelasan mengenai validitas logis dan validitas empiris dibahas pada halaman berikut.

a. Validitas Logis

1. Validitas isi (*content validity*). Sebuah tes dikatakan memiliki validitas isi apabila mengukur tujuan khusus tertentu yang sejajar dengan materi atau isi pelajaran yang diberikan. Menurut Sudaryono (2012:141) validitas isi suatu tes hasil belajar dapat diketahui dengan membandingkan antara isi yang terkandung dalam tes hasil belajar, dengan tujuan instruksional khusus pelajaran tersebut atau dalam istilah yang dipakai saat ini yaitu kompetensi dasar.
2. Validitas konstruksi (*construct validity*). Sebuah tes dikatakan memiliki validitas konstruksi apabila butir-butir yang membangun tes tersebut mengukur setiap aspek berpikir seperti yang disebutkan dalam Tujuan Instruksional Khusus.

b. Validitas Empiris

1. Validitas “ada sekarang” (*concurrent validity*). Validitas ini lebih umum dikenal dengan validitas empiris. Sebuah tes dikatakan memiliki validitas empiris jika hasilnya sesuai dengan pengalaman.
2. Validitas prediksi (*predictive validity*). Sebuah tes dikatakan memiliki validitas prediksi atau validitas ramalan apabila tes mempunyai kemampuan untuk meramalkan apa yang akan terjadi pada masa yang akan datang.

Dalam penelitian ini analisis validitas logis yang dilakukan adalah analisis validitas isi pada UAS, yaitu dengan membandingkan materi yang terdapat pada UAS dengan Peta Uraian Materi. Sementara pada validitas empiris analisis dilakukan pada data statistik dari hasil jawaban UAS tersebut.

Sebuah tes dikatakan memiliki validitas jika hasilnya sesuai dengan kriteria, dalam arti memiliki kesejajaran antara hasil tersebut dengan kriteria. Artinya skor total suatu butir berkorelasi dengan skor total tes tersebut. Untuk menentukan sebuah apakah sebuah tes sudah memenuhi kesesejajaran atau korelasi tersebut, digunakan teknik korelasi *product moment* yang dikemukakan oleh Pearson dalam Arikunto (2006:75), dengan interpretasi besarnya koefisien korelasi seperti berikut:

Antara 0,800 sampai dengan 1,00 : 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

A.3.2 Reliabilitas

Kata reliabilitas dalam bahasa Indonesia diambil dari kata *reliability* dalam bahasa Inggris, berasal dari kata *reliable* yang artinya dapat dipercaya. Arikunto (2006:60) menyatakan bahwa sebuah tes dikatakan reliabel apabila hasil tes tersebut menunjukkan ketetapan. Senada dengan Arikunto, Sukardi (2009:43) menyatakan bahwa instrumen evaluasi mempunyai reliabilitas tinggi apabila tes mempunyai hasil yang konsisten yang membuat terpenuhinya syarat utama, yaitu validnya suatu hasil skor instrumen. Senada dengan Arikunto dan Sukardi, Bausch, Christ dan Krumm (2007:366) menyatakan “*Das Kriterium der Reliabilität dagegen bezieht sich auf die Reproduzierbarkeit von Testergebnissen innerhalb bestimmter Toleranzgrenzen (Grad der Genauigkeit des Messvorgangs).*” Disebutkan bahwa kriteria reliabilitas berhubungan dengan kemampuan tes untuk memberikan hasil yang sama atau memberikan ketepatan dari pengukuran. Dengan demikian dapat disimpulkan bahwa sebuah tes dikatakan reliabel apabila ketika diujikan berkali-kali akan memberikan hasil yang sama, atau apabila terjadi perubahan, tidak terdapat perubahan yang berarti.

Arikunto (2006:90) menyatakan besarnya reliabilitas sebuah tes dapat diketahui cara-cara seperti bentuk paralel, metode tes ulang dan metode belah dua:

- a. Metode bentuk paralel (*equivalent*). Dalam metode ini pengetes harus menyiapkan dua buah tes, dan masing-masing dicobakan pada kelompok siswa yang sama.

- b. Metode tes ulang (*test-retest method*). Dalam metode ini pengetes hanya memiliki satu seri tes namun dicobakan dua kali.
- c. Metode belah dua atau *split-half method*. Dalam metode ini pengetes hanya mempergunakan sebuah tes yang diujicobakan sekali.

Selain tiga cara di atas Surapranata (2009:113) menambahkan satu jenis metode untuk mencari reliabilitas yaitu dengan *internal consistency*. Metode *internal consistency* didasarkan pada homogenitas atau korelasi antar skor jawaban pada setiap butir tes. Teknik yang digunakan dengan metode ini adalah teknik koefisien alpha. Pada penelitian ini tingkat reliabilitas akan dianalisis melalui teknik koefisien alpha tersebut.

Koefisien reliabilitas dilambangkan dengan r_{11} dan bergerak dari angka 0.0 sampai 1.0. Artinya apabila sebuah tes memiliki tingkat reliabilitas yang baik apabila koefisien reliabilitas tes tersebut makin mendekati angka 1.0. Sementara itu menurut Sudijono (2009:209), pada koefisien reliabilitas umumnya digunakan patokan sebagai berikut:

1. Apabila r_{11} sama dengan atau lebih besar daripada 0,70 berarti tes hasil belajar yang sedang diuji reliabilitasnya dinyatakan telah memiliki reliabilitas yang tinggi (*=reliable*).
2. Apabila r_{11} lebih kecil dari pada 0,70 berarti bahwa tes hasil belajar yang sedang diuji reliabilitasnya dinyatakan belum memiliki reliabilitas yang tinggi (*unreliable*)

A.3.3 Tingkat Kesukaran

Tingkat kesukaran adalah pengukuran seberapa sukar sebuah soal dilihat dari pilihan jawaban siswa. Menurut Gruber dan Avallone (2012:26), "*Item-Schwierigkeit erfasst den Anteil der Prüfungsteilnehmer, die eine Frage richtig beantwortet haben.*" Disebutkan bahwa tingkat kesukaran butir adalah proporsi jumlah peserta tes yang menjawab suatu pertanyaan dengan benar. Hal yang sama dikemukakan oleh Sukardi (2009:136) yang menyatakan bahwa tingkat kesulitan item atau indeks kesulitan item adalah angka mengenai proporsi siswa yang menjawab betul pada satu butir soal dalam tes objektif. Angka yang menunjukkan mengenai tingkat kesukaran dikenal dengan "*Difficulty Index*" yang diberi lambang P (*Proportion*). Besarnya tingkat kesukaran antara 0,00 sampai dengan 1,0 yang berarti soal dengan indeks kesukaran 0,00 menunjukkan bahwa soal tersebut terlalu sukar, sebaliknya jika indeks kesukaran 1,0 maka soal tersebut terlalu mudah. Indeks kesukaran biasa diberi simbol P (singkatan dari kata proporsi). Menurut Arikunto (2006:210), klasifikasi indeks kesukaran adalah sebagai berikut:

Soal dengan P 1,00 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.

A.3.4 Daya Pembeda

Arikunto (2006:211) menyatakan bahwa daya pembeda adalah kemampuan suatu soal untuk membedakan antara siswa pandai dengan siswa

bodoh. Apabila suatu soal yang dapat dijawab dengan benar oleh siswa pandai maupun bodoh, maka soal tersebut dapat disebut tidak memiliki daya pembeda. Begitu juga apabila suatu soal tidak dapat dijawab dengan benar oleh siswa pandai maupun bodoh, maka dan soal tersebut juga dapat disebut tidak memiliki daya pembeda.

Angka yang menunjukkan besarnya daya pembeda adalah disebut indeks diskriminasi yang diberi lambang D . Besarnya daya pembeda berkisar antara 0,00 sampai 1,00. Pada indeks diskriminasi juga dikenal tanda negatif. Tanda negatif pada indeks diskriminasi digunakan apabila suatu butir soal “terbalik” menunjukkan kualitas *testee* (peserta tes). Artinya siswa pandai tidak dapat menjawab butir tersebut sedangkan siswa bodoh dapat menjawab butir tersebut.

Menurut Arikunto (2006:218) klasifikasi daya pembeda soal adalah sebagai berikut:

$D = 0,00 - 0,20$ maka daya pembeda soal adalah jelek (*poor*)

$D = 0,21 - 0,40$ maka daya pembeda soal adalah cukup (*satisfactory*)

$D = 0,41 - 0,70$ maka daya pembeda soal adalah baik (*good*)

$D = 0,71 - 1,00$ maka daya pembeda soal adalah baik sekali (*excellent*)

$D = \text{negatif}$, maka semuanya tidak baik, jadi semua butir soal yang mempunyai nilai D negatif sebaiknya dibuang saja.

Jadi dapat disimpulkan bahwa daya pembeda adalah kemampuan suatu butir soal dalam membedakan siswa berkemampuan tinggi (siswa pandai) dan siswa berkemampuan rendah (siswa bodoh).

A.3.5 Fungsi Pengecoh

Pada sebuah soal pilihan ganda selain jawaban kunci (jawaban yang benar) terdapat alternatif jawaban/*option* yang merupakan pengecoh (distraktor). Distraktor berfungsi untuk mengecoh siswa saat memilih jawaban yang paling tepat yang tersedia pada *options*.

Dalam E-Learning Bauskasten (http://www.let.ethz.ch/pruefungen/umsetzung/dienstleistungen/Leitfaden_MCfragen.pdf, diakses tanggal 29 Desember 2013), “*Distraktoren unterscheiden sich signifikant von der richtigen Antwort und schließen sich gegenseitig aus.*” Disebutkan bahwa pengecoh secara signifikan berbeda dan saling bertolak belakang dengan jawaban yang tepat. Sementara itu Arikunto (2006:220) menjelaskan bahwa pengecoh (distraktor) yang sama sekali tidak dipilih oleh *testee* menandakan bahwa pengecoh itu jelek, menyolok dan menyesatkan. Oleh karena itu distraktor dalam soal pilihan ganda harus mempunyai fungsi yang baik dalam mengecoh siswa. Suatu distraktor dapat dikatakan berfungsi apabila distraktor tersebut dipilih oleh paling sedikit 5% dari jumlah pengikut tes.

B. Penelitian Yang Relevan

Beberapa penelitian yang relevan dengan penelitian analisis butir soal ini telah ditemukan diantaranya :

1. Vonny Muliawati, Analisis Butir Soal Pada Ulangan Blok Semester 1 Bahasa Jerman di SMA Kelas XII, Jakarta : Jurusan Pendidikan Bahasa Jerman, FBS, UNJ, 2010.

2. Kartini Megawati, Kualitas Soal Ulangan Umum Semester 1 Bahasa Jerman Kelas X SMAN 48 Jakarta Tahun Pelajaran 2005/2006, Jakarta : Jurusan Pendidikan Bahasa Jerman, FBS, UNJ, 2007.

Dari dua penelitian yang telah disebutkan, terdapat persamaan dengan penelitian ini, yaitu menganalisis validitas butir, reliabilitas, kesukaran, dan daya pembeda serta fungsi pengecoh yang terdapat pada soal ujian atau ulangan. Sedangkan perbedaannya adalah pada penelitian ini dianalisis validitas isi dari soal ujian, yaitu dengan cara membandingkan materi yang terdapat pada soal dengan materi yang seharusnya diajarkan pada siswa di kelas.

C. Kerangka Berpikir

Mata pelajaran bahasa Jerman di Sekolah Menengah (SMA) disesuaikan dengan Kurikulum Tingkat Satuan Pendidikan (KTSP). Pemberlakuan kurikulum ini memberi ruang pada guru untuk menyusun evaluasi belajar. Artinya guru dapat membuat soal ujian sendiri dengan tetap berpegangan pada standar kompetensi dan kompetensi dasar yang berlaku. Evaluasi pada dasarnya merupakan kegiatan yang wajib dilakukan pada semua jenjang pendidikan karena dengan evaluasi guru dan siswa untuk mengetahui seberapa baik siswa menguasai materi yang sudah diajarkan oleh guru dalam kurun waktu tertentu.

Kegiatan evaluasi tidak bisa lepas dari salah satu bentuk evaluasi itu sendiri yaitu tes. Terkadang sebuah tes tidak berkualitas baik sehingga tidak dapat dengan dapat mengukur hasil belajar siswa, padahal dalam menyusun sebuah tes,

guru harus memperhatikan kriteria-kriteria tes yang baik yang terdiri dari kriteria valid, reliabel, obyektif dan praktis. Artinya tes tersebut harus dapat mengukur apa yang harus diukur, dapat dipercaya atau dapat memberikan hasil yang sama ketika diteskan berkali-kali, tidak terdapat unsur subyektifitas dalam proses penilaiannya dan mudah untuk dilaksanakan dan diperiksa.

Tes obyektif dengan soal berbentuk pilihan ganda digunakan untuk tes dengan materi yang banyak seperti contohnya ujian akhir semester (UAS). Hal ini karena materi yang ada pada satu semester sangat banyak. Misalnya pada pelajaran bahasa Jerman kelas 12 semester 1, diajarkan materi bertema Hobi/Kegemaran. Materi ini selain untuk melatih keterampilan berbahasa seperti mendengarkan, berbicara, membaca dan menulis, juga untuk melatih *Redemittel*, *Struktur* dan *Wortschatzt*. Tes bentuk obyektif juga memudahkan guru dalam proses penilaian karena cukup menggunakan kunci jawaban yang telah dipersiapkan sebelumnya sehingga penilaian tes ini bersifat obyektif.

Untuk mengetahui kualitas sebuah tes perlu dilakukan analisis pada butir-butir soal tersebut. Analisis ini dikenal dengan analisis butir soal. Pada penelitian ini peneliti menggunakan analisis kuantitatif untuk menghitung validitas, reliabilitas, tingkat kesukaran dan daya pembeda serta fungsi pengecoh. Peneliti juga melakukan analisis validitas isi pada soal. Pada penelitian ini dianalisis butir soal ujian akhir bahasa Jerman kelas XII tahun ajaran 2012/2013 semester 1 milik SMA Negeri 54 Jakarta.

Tingkat validitas dicari dengan menggunakan teknik korelasi *product moment* Pearson. Sedangkan validitas isi dapat diketahui dengan cara

membandingkan materi yang terdapat pada soal ujian dengan materi yang terdapat pada kurikulum. Hal ini sangat penting karena terkadang materi dalam soal ujian tidak termasuk dalam kurikulum di kelas dan semester tersebut.

Reliabilitas dianalisis untuk mengetahui tingkat keajegan hasil tes atau hasil sebuah tes ketika diujikan berkali-kali. Analisis reliabilitas dilakukan dengan menggunakan teknik koefisien alpha. Analisis tingkat kesukaran dilakukan untuk menentukan butir-butir soal yang sukar, sedang dan mudah. Analisis daya pembeda dilakukan untuk mengidentifikasi butir-butir soal mana saja yang dapat membedakan siswa berkemampuan tinggi dan siswa berkemampuan rendah. Sebuah dengan daya pembeda baik dapat dijawab dengan benar oleh siswa berkemampuan tinggi dan dijawab dengan salah oleh siswa berkemampuan rendah. Analisis fungsi pengecoh dilakukan untuk mengetahui seberapa efektif distraktor-distraktor pada butir soal pilihan ganda dalam mengecoh siswa pada saat memilih jawaban yang benar. Hasil analisis dan perhitungan akan dimasukkan dalam tabel analisis. Kemudian hasil tersebut diinterpretasikan dan ditarik kesimpulan. Setelah itu akan diketahui butir-butir soal mana saja yang berkualitas dalam ujian akhir semester bahasa Jerman kelas XII tahun ajaran 2012/2013 semester 1 di SMA Negeri 54 Jakarta.

Dalam penelitian ini digunakan soal ujian akhir semester bahasa Jerman kelas XII tahun ajaran 2012/2013 semester 1 dan laporan perorangan UAS yang didapatkan peneliti dari SMA Negeri 54 Jakarta sebagai sumber data dalam melakukan analisis butir soal.

D. Definisi Konseptual

Proses belajar yang dilakukan pada satu semester memerlukan evaluasi untuk melihat kemajuan belajar siswa. Ujian Akhir Semester (UAS) merupakan salah satu bentuk evaluasi yang berupa tes yang diberikan kepada siswa pada akhir semester. Untuk mengetahui kualitas setiap butir soal tes tersebut diperlukan analisis butir soal. Analisis butir soal dilakukan untuk mengetahui tingkat validitas yang berupa validitas (isi dan empiris), reliabilitas, kesukaran, daya pembeda serta fungsi pengecoh.

Pada penelitian ini validitas yang dianalisis adalah validitas isi dan empiris. Validitas isi dianalisis menggunakan peta uraian materi bahasa Jerman kelas XII semester 1. Sementara validitas empiris dianalisis menggunakan korelasi Pearson sehingga diketahui koefisien validitas. Koefisien untuk tingkat validitas adalah antara 0,800 sampai dengan 1,00: sangat tinggi, antara 0,600 sampai dengan 0,800: tinggi, antara 0,400 sampai dengan 0,600: cukup, antara 0,200 sampai dengan 0,200: rendah, antara 0,00 sampai dengan 0, 200: sangat rendah.

Realibilitas adalah ukuran keajegan hasil sebuah tes ketika diujikan berkali-kali. Reabilitas dicari dengan menggunakan rumus alpha. Koefisien reliabilitas tes (r_{11}) ditunjukkan dengan apabila r_{11} lebih tinggi dari 0.8 maka tes tersebut sudah termasuk reliabel.

Tingkat kesukaran adalah ukuran mengenai seberapa sukar atau mudah butir soal. Cara untuk mengetahui tingkat kesukaran adalah dengan membagi jumlah siswa yang menjawab benar sebuah butir soal dengan jumlah seluruh

siswa yang mengikuti ujian. Indeks untuk tingkat kesukaran (P) dibagi menjadi 3 yaitu apabila P 1,00 sampai 0,30 maka soal tersebut sukar, apabila P 0,30 sampai 0,70 maka soal tersebut sedang dan apabila P 0,70 sampai 1,00 maka soal tersebut mudah.

Daya pembeda adalah kemampuan suatu soal untuk membedakan siswa yang berkemampuan tinggi dan siswa yang berkemampuan rendah. Dalam menghitung daya pembeda, skor seluruh siswa peserta ujian harus diurutkan mulai dari tertinggi sampai terendah. Proporsi untuk menentukan siswa berkemampuan tinggi adalah 27% siswa dengan skor tertinggi dan untuk menentukan siswa berkemampuan rendah adalah 27% siswa dengan skor terendah. Setelah menentukan siswa berkemampuan tinggi dan rendah, daya pembeda dihitung dengan cara membagi mengurangi proporsi peserta berkemampuan tinggi yang menjawab soal dengan benar dengan proporsi peserta berkemampuan rendah yang menjawab soal dengan benar. Ukuran untuk menentukan seberapa baik daya pembeda soal disebut indeks diskriminasi (D) dengan klasifikasi apabila $D = 0,00 - 0,20$ maka daya pembeda soal adalah jelek, $D = 0,21 - 0,40$ daya pembeda soal adalah cukup, $D = 0,41 - 0,70$ daya pembeda soal adalah baik (good), $D = 0,71 - 1,00$ daya pembeda soal adalah baik sekali, dan apabila $D = \text{negatif}$, maka soal tidak baik dan sebaiknya dibuang saja.

Fungsi pengecoh adalah kemampuan sebuah distraktor untuk mengecoh peserta tes. Sebuah tes dikatakan memiliki fungsi pengecoh yang baik apabila distraktor dipilih oleh minimal 5% dari siswa peserta tes.

E. Definisi Operasional

Untuk menghindari penafsiran yang berbeda, diperlukan penjelasan mengenai istilah yang digunakan pada penelitian ini melalui definisi operasional sebagai berikut:

1. Analisis butir soal yang dimaksud dalam penelitian ini adalah analisis yang dilakukan pada butir soal ujian akhir semester bahasa Jerman kelas XII tahun ajaran 2012/2013 semester 1 milik SMA Negeri 54 Jakarta. Pada penelitian ini analisis butir soal yang dilakukan meliputi hal sebagai berikut yaitu tingkat validitas, tingkat reliabilitas, tingkat kesukaran, dan daya pembeda serta fungsi pengecoh.
2. Ujian akhir semester bahasa Jerman kelas XII tahun ajaran 2012/2013 semester 1 yang dimaksud dalam penelitian ini adalah ujian akhir semester mata pelajaran bahasa Jerman milik SMA Negeri 54 yang dikerjakan oleh seluruh siswa kelas XII tahun ajaran 2012/2013 semester 1 pada tanggal 4 Desember 2012.