

BAB I PENDAHULUAN

1.1. Latar Belakang Masalah

Evaluasi merupakan proses yang digunakan di berbagai bidang untuk menilai pencapaian tujuan dan mengidentifikasi hal-hal yang perlu diperbaiki dalam kehidupan sehari-hari, evaluasi membantu pengambilan keputusan yang lebih baik dan mendorong perbaikan berkelanjutan. Dalam pendidikan, evaluasi berperan penting untuk menilai apakah proses pembelajaran telah berjalan sesuai dengan tujuan yang ditetapkan. Pada lingkup terbatas, evaluasi menilai pencapaian belajar peserta didik secara individu. Sedangkan pada lingkup luas, evaluasi digunakan untuk menilai keberhasilan dan kelemahan proses pembelajaran dalam mencapai tujuan pendidikan.

Sebagai calon pendidik, evaluasi pembelajaran berperan penting dalam mendukung pengambilan keputusan yang tepat demi kemajuan peserta didik secara individu maupun peningkatan kualitas pendidikan secara keseluruhan. Setiap langkah dalam proses evaluasi selalu diarahkan pada pencapaian hasil yang mencerminkan efektivitas pembelajaran. Pendidik selalu berharap bahwa hasil yang diperoleh sekarang lebih baik dan memuaskan dari hasil yang diperoleh sebelumnya, untuk menentukan dan membandingkan antara satu hasil dengan lainnya diperlukan adanya evaluasi pembelajaran (Ratnawulan and Rusdiana 2014:13–14). Secara umum, evaluasi pembelajaran bertujuan untuk mendapatkan bukti dan data yang dapat menunjukkan seberapa efektif proses belajar yang telah dilakukan. Tujuannya juga untuk mengetahui sejauh mana siswa telah memahami materi sesuai dengan kurikulum yang berlaku. Dalam hal ini, agar hasil evaluasi akurat dan tepat sasaran, diperlukan alat evaluasi yang sesuai dengan tujuan pembelajaran, serta indikator yang telah ditentukan sebagai dasar dalam menilai kemajuan siswa (Otaya et al. 2023:40). Istilah lain yang erat hubungannya dengan evaluasi dan pengukuran adalah penilaian (*assessment*). Dalam sistem pendidikan, terdapat berbagai jenis asesmen yang digunakan untuk mengukur capaian pembelajaran peserta didik, salah satunya adalah asesmen sumatif.

Asesmen sumatif adalah penilaian yang dilakukan setelah selesai mengajar sebuah program pelajaran. Dengan kata lain, asesmen ini dilakukan setelah seluruh materi pelajaran telah diajarkan. Tujuan utama dari asesmen sumatif adalah menentukan nilai yang mencerminkan keberhasilan siswa setelah mereka menyelesaikan program pengajaran dalam jangka waktu tertentu (Maharani 2018:20). Jadi, asesmen sumatif merupakan penilaian yang dilaksanakan pada akhir unit program, misalnya pada akhir caturwulan, akhir semester, atau akhir tahun. Penilaian ini berorientasi pada hasil belajar siswa dan dapat menentukan status keberhasilan peserta didik dalam program tertentu. Contohnya meliputi tes caturwulan, tes akhir semester, dan Evaluasi Belajar Tahap Akhir (EBTA) (Maharani 2018:20). Dalam konteks pendidikan kejuruan, penilaian selayaknya mampu memberikan gambaran yang akurat mengenai keterampilan yang telah dikuasai siswa sesuai dengan kebutuhan industri. Oleh karena itu, sistem evaluasi yang diterapkan di sekolah harus selaras dengan pendekatan kurikulum yang digunakan.

Kurikulum Merdeka yang diterapkan di SMK saat ini menekankan pada pembelajaran yang fleksibel dan berbasis proyek. Evaluasi dalam Kurikulum Merdeka harus mampu mengukur kompetensi siswa secara menyeluruh, baik dari aspek akademik maupun keterampilan yang relevan dengan dunia kerja. Oleh karena itu, Penilaian (*assessment*) adalah proses pengumpulan dan pengolahan informasi untuk mengukur pencapaian hasil belajar siswa yang bertujuan untuk membantu kemajuan dan pencapaian tujuan pembelajaran yang dilaksanakan peserta didik (Maharani 2018:22). Dalam konteks pembelajaran di SMK Negeri Jakarta, mata pelajaran Instalasi Tenaga Listrik merupakan mata pelajaran wajib bagi siswa kelas XI di program keahlian Teknik Instalasi Tenaga Listrik. Mata pelajaran ini tidak hanya bertujuan untuk memberikan pemahaman konsep dasar kelistrikan, tetapi juga melatih siswa agar memiliki keterampilan teknis dalam pemasangan, perawatan, dan perbaikan instalasi listrik.

Sebagai fasilitator dalam pembelajaran, guru memiliki peran penting dalam menyediakan soal ujian yang sesuai dengan standar evaluasi yang baik. Guru yang kompeten tidak hanya dituntut mampu menyampaikan materi secara efektif, tetapi juga harus memiliki keterampilan dalam menyusun alat ukur yang mampu

mengukur tingkat pemahaman dan pencapaian kompetensi siswa secara akurat. Sayangnya, di lapangan masih ditemukan kecenderungan guru menggunakan soal yang disusun tanpa melalui proses validasi atau analisis butir soal terlebih dahulu. Soal-soal tersebut langsung diujikan kepada peserta didik tanpa diketahui terlebih dahulu kualitasnya secara empiris, baik dari segi validitas, reliabilitas, daya beda, maupun tingkat kesukaran. Akibatnya, instrumen yang digunakan dalam asesmen sering kali tidak mencerminkan perkembangan kurikulum maupun karakteristik siswa secara aktual.

Kondisi tersebut menyebabkan ketimpangan distribusi tingkat kesukaran soal. Soal menjadi terlalu mudah bagi siswa dengan kemampuan tinggi dan terlalu sulit bagi siswa yang pemahamannya masih rendah. Hal ini berdampak pada hasil evaluasi yang tidak mencerminkan kemampuan siswa secara objektif. Dengan kata lain, proses evaluasi kehilangan fungsinya sebagai alat pengukur yang adil dan akurat. Beberapa siswa mengalami kesulitan menjawab soal bukan karena tidak memahami materi, tetapi karena tingkat kesukarannya tidak sesuai dengan profil kemampuan mereka. Soal yang terlalu sulit bisa membuat sebagian siswa mengalami kesulitan dalam memahami dan menyelesaikannya, sehingga mengganggu fokus mereka saat mengerjakan. Sebaliknya, soal yang terlalu mudah untuk siswa dengan kemampuan tinggi justru kurang menantang dan tidak mendorong mereka untuk berpikir lebih kritis. Akibatnya, proses evaluasi jadi kurang adil dan tidak benar-benar menggambarkan kemampuan siswa secara menyeluruh .

Sejumlah penelitian telah mengungkapkan bahwa penyebab kesalahan dalam asesmen bukan semata-mata berasal dari peserta didik, melainkan dari kualitas soal yang belum memenuhi kriteria instrumen evaluasi yang baik. Misalnya, penelitian oleh Alti et al. (2023) menyebutkan bahwa hasil analisis butir soal pilihan ganda Ujian Tengah Semester mata pelajaran Biologi Kelas X di MAN 1 Solok Selatan menunjukkan bahwa soal-soal yang digunakan belum memenuhi standar mutu yang diharapkan. Hal ini terlihat dari validitas soal yang hanya lima butir tergolong sangat valid, reliabilitasnya rendah (0,37), distribusi daya pembeda yang didominasi kategori cukup hingga jelek, serta tingkat kesukaran yang belum

merata. Oleh karena itu, disimpulkan bahwa soal-soal tersebut belum layak digunakan untuk mengukur ketercapaian pembelajaran secara akurat.

Penelitian serupa juga diungkapkan oleh Amiruddin et al. (2020) yang menganalisis soal Ujian Akhir Sekolah mata pelajaran matematika di MTs.S Guppi Bongtolangkasa. Dari hasil analisis diperoleh bahwa sebagian besar soal berada pada kategori mudah (70%), dengan daya pembeda yang rendah, di mana hanya 5% soal yang diterima, dan 70% soal masuk dalam kategori ditolak. Efektivitas pengecoh pun menunjukkan bahwa sebagian besar opsi jawaban tidak berfungsi secara optimal. Berdasarkan ketiga indikator kualitas soal, yakni tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Dapat disimpulkan bahwa sebagian besar soal tidak memenuhi kriteria sebagai butir soal yang baik.

Berdasarkan dua temuan tersebut, terlihat bahwa permasalahan dalam evaluasi pembelajaran lebih banyak disebabkan oleh instrumen soal yang belum dirancang dan dianalisis secara memadai, sehingga belum dapat mencerminkan pencapaian kompetensi siswa secara objektif dan belum memberikan gambaran yang utuh terhadap ketercapaian Tujuan Pembelajaran (TP) maupun Capaian Pembelajaran (CP) sesuai struktur kurikulum yang berlaku. Dengan kata lain, ketidaksesuaian antara karakteristik soal dengan kemampuan peserta didik menjadi faktor utama yang mempengaruhi rendahnya kualitas hasil evaluasi, bukan semata-mata karena keterbatasan siswa itu sendiri.

Temuan-temuan tersebut menegaskan bahwa kualitas soal yang baik tidak hanya tergantung pada isi materi, tetapi juga pada proporsionalitas tingkat kesukarannya terhadap kemampuan peserta tes. Oleh karena itu, diperlukan analisis butir soal secara sistematis guna mengevaluasi validitas, reliabilitas, daya beda, efektivitas pengecoh, dan tingkat kesukaran setiap *Item*. Dengan analisis ini, guru dapat mengetahui sejauh mana instrumen yang digunakan telah memenuhi kriteria sebagai alat ukur yang baik. Selain itu, analisis butir soal memberikan dasar yang kuat bagi guru dalam melakukan revisi terhadap soal-soal yang tidak layak, serta menyusun soal baru yang sesuai dengan kebutuhan asesmen pembelajaran yang objektif dan berorientasi pada kemampuan siswa, serta sejauh mana soal-soal tersebut mendukung pencapaian TP yang terintegrasi dalam kurikulum melalui CP sebagai acuan kompetensi utama. Selain itu, Evaluasi pembelajaran dalam

Kurikulum Merdeka juga harus mencerminkan ketercapaian TP dan CP, sebagai pedoman kompetensi dan arah asesmen yang terukur. Oleh karena itu Soal yang dikembangkan perlu mencerminkan ketercapaian TP secara praktis, yang dalam penelitian ini diidentifikasi melalui proporsi siswa yang menjawab benar pada soal yang dipetakan ke TP. Dengan pendekatan ini, evaluasi tidak hanya berfokus pada validitas dan reliabilitas soal, tetapi juga pada keterhubungan antara hasil belajar peserta didik dan tujuan pembelajaran yang telah ditetapkan. Hal inilah yang menjadi dasar munculnya beberapa masalah utama yang akan diteliti, seperti belum dianalisisnya soal-soal buatan guru secara sistematis, penggunaan soal yang berulang tanpa validasi, hingga ketidaksesuaian tingkat kesukaran soal dengan kemampuan siswa.

Penelitian ini bertujuan untuk menganalisis kualitas teknis butir soal menggunakan pendekatan Model Rasch, serta mengevaluasi ketercapaian Tujuan Pembelajaran (TP) sebagai turunan dari Capaian Pembelajaran (CP) melalui pemanfaatan data persentase siswa yang menjawab benar pada setiap soal berdasarkan *output Ministep*. Selain menilai kualitas teknis butir soal, penelitian ini juga diarahkan untuk mengevaluasi ketercapaian Tujuan Pembelajaran (TP) yang merupakan turunan dari Capaian Pembelajaran (CP) dalam Kurikulum Merdeka. TP berfungsi sebagai acuan pembelajaran yang konkret di kelas, sehingga keberhasilan peserta didik dalam menjawab soal yang mewakili TP dapat mencerminkan sejauh mana CP telah tercapai. Penelitian difokuskan pada SMK Negeri 5 dan SMK Negeri 26 Jakarta, yang berdasarkan hasil konfirmasi dari guru produktif dan pihak sekolah, soal ujian yang digunakan sebelumnya belum pernah dianalisis secara sistematis dengan pendekatan Model Rasch. Melalui model ini, dapat diidentifikasi butir soal yang tergolong baik dan yang perlu direvisi atau dihilangkan. Dengan demikian, instrumen yang digunakan benar-benar dapat mengukur hasil belajar kognitif siswa secara objektif, terukur, serta mendukung ketercapaian TP dan CP sesuai dengan prinsip evaluasi pembelajaran yang bermutu.

Analisis butir soal dilakukan dengan menggunakan Model Rasch, yang merupakan bagian dari Teori Respon Butir (*Item Response Theory//IRT*). Model Rasch memiliki keunggulan dibandingkan dengan Teori Tes Klasik (*Classical Test Theory/CTT*), karena mampu mengevaluasi kualitas soal tanpa terpengaruh oleh

karakteristik peserta tes. Pemodelan Rasch pada penilaian pendidikan memberikan pendekatan yang berbeda pada penggunaan skor atau data mentah. Tidak seperti pendekatan klasik yang sangat tergantung pada rata-rata dan distribusi skor, Model Rasch mengubah data mentah menjadi skala logit yang bersifat interval, artinya jarak antar nilai sama besar sehingga lebih tepat untuk analisis kuantitatif. Dengan pendekatan ini, diperoleh informasi yang lebih mendalam baik mengenai karakteristik peserta tes maupun kualitas butir soal yang dikerjakan. Hal ini dimungkinkan karena Model Rasch membentuk suatu rasio pengukuran dengan jarak antar nilai yang sama, sehingga interpretasi data menjadi lebih objektif dan bermakna (Purniasari, Masykuri, and Ariani 2021).

Dengan keunggulan tersebut tepat jika pendekatan *IRT*, khususnya Model Rasch digunakan untuk menganalisis kualitas soal ujian. Melalui analisis ini, dapat diketahui apakah suatu butir soal memiliki tingkat kesulitan yang sesuai, daya pembeda yang optimal, dan distraktor (pengecoh) yang efektif. Informasi ini sangat penting untuk mengetahui butir soal mana yang berkualitas baik dan layak digunakan, serta butir soal mana yang perlu direvisi atau bahkan dibuang. Dengan menggunakan Model Rasch, hasil analisis butir soal dapat memberikan gambaran yang lebih akurat mengenai kualitas instrumen evaluasi. Hasil analisis ini menjadi dasar untuk meningkatkan kualitas soal, yang pada akhirnya berdampak pada peningkatan mutu pendidikan. Mutu pendidikan yang baik tidak hanya diukur dari hasil ujian siswa, tetapi juga dari efektivitas metode pengajaran, kualitas soal evaluasi, serta kemampuan guru dalam menyusun asesmen yang objektif dan komprehensif. Dengan adanya analisis butir soal, diharapkan evaluasi yang dilakukan dapat lebih akurat, sehingga siswa dapat memperoleh pengalaman belajar yang lebih optimal dan sesuai dengan standar kompetensi yang diharapkan. Dengan demikian, pendekatan ini tidak hanya bermanfaat secara teknis tetapi juga penting dalam memastikan bahwa asesmen mendukung ketercapaian TP dan CP secara praktis di lapangan.

State of the Art Penelitian terkait analisis butir soal telah banyak dilakukan sebelumnya baik di jenjang SMP, SMA, maupun SMK, dengan menggunakan pendekatan teori tes klasik (*CTT*) maupun teori respon butir (*IRT*), termasuk Model Rasch. Beberapa penelitian terdahulu menunjukkan bahwa masih banyak soal

evaluasi yang belum memenuhi standar kualitas yang baik, baik dari segi validitas, reliabilitas, daya pembeda, hingga efektivitas pengecoh. Misalnya, penelitian oleh Qurrotul Ayun dan Pantiwati (2024) menemukan bahwa meskipun reliabilitas soal cukup tinggi (0,72), masih ada soal yang berkategori tidak valid dan pengecoh yang tidak berfungsi optimal. Di sisi lain, Erfan et al. (2020) menunjukkan adanya perbedaan hasil antara pendekatan klasik dan Rasch dalam menganalisis soal IPA listrik, di mana dari 40 soal, hanya 5 yang valid menurut Rasch meskipun dalam *CTT* tergolong baik.

Sebagian besar penelitian tersebut fokus pada mata pelajaran umum dan belum banyak menyorot pada konteks pendidikan vokasi, khususnya mata pelajaran produktif seperti Instalasi Tenaga Listrik di SMK. Selain itu, pendekatan yang digunakan juga masih terbatas, dengan beberapa penelitian yang belum mengintegrasikan secara komprehensif kelima indikator utama kualitas soal: validitas, reliabilitas, tingkat kesukaran, daya beda, dan efektivitas pengecoh, dalam satu model analisis. Selain itu, belum banyak penelitian yang secara eksplisit mengaitkan hasil analisis soal dengan ketercapaian Tujuan Pembelajaran (TP) dan Capaian Pembelajaran (CP) dalam Kurikulum Merdeka.

Kebaruan (*novelty*) dalam penelitian ini terletak pada penerapan Model Rasch berbasis aplikasi *Ministep* untuk menganalisis soal Ujian Akhir Semester mata pelajaran Instalasi Tenaga Listrik kelas XI di dua SMK Negeri di Jakarta. Sekolah vokasi seperti ini masih jarang diteliti, terutama dalam hal analisis kualitas soal. Penelitian ini menjadi penting karena fokusnya pada soal buatan guru mata pelajaran produktif yang sebelumnya belum pernah dianalisis secara mendalam, baik dari segi kualitas teknis maupun sejauh mana soal tersebut mendukung ketercapaian Tujuan Pembelajaran (TP) sebagai turunan dari Capaian Pembelajaran (CP). Hasil dari penelitian ini tidak hanya berguna untuk membantu guru menyusun bank soal yang lebih baik, tapi juga mendorong guru agar lebih reflektif dan berbasis data dalam menyusun soal evaluasi di pendidikan kejuruan.

Dengan demikian, penelitian ini tidak hanya mengungkap kelemahan soal yang digunakan, tetapi juga memberikan rekomendasi konkret untuk penyusunan soal yang lebih berkualitas dan sesuai dengan kebutuhan pembelajaran di SMK. Lebih lanjut, hasil analisis butir soal ini dapat digunakan sebagai dasar dalam

pengelolaan bank soal. Soal yang memiliki validitas yang baik dan memenuhi standar evaluasi dapat disimpan dan digunakan kembali, sementara soal yang memiliki kualitas rendah dapat direvisi atau dihilangkan dari bank soal. Dengan sistem tersebut, kualitas evaluasi dapat ditingkatkan secara berkelanjutan dan guru memiliki referensi soal yang lebih terstandarisasi untuk asesmen berikutnya. Hasil penelitian ini diharapkan dapat menjadi acuan bagi guru dalam menyusun soal evaluasi yang lebih efektif, yang tidak hanya layak secara teknis tetapi juga mendukung ketercapaian TP dan CP sesuai dengan arah kurikulum yang berlaku.

1.2. Identifikasi Masalah

Berdasarkan latar belakang yang telah dijelaskan, terdapat beberapa permasalahan utama yang dapat diidentifikasi dalam penelitian ini, yaitu:

1. Soal-soal yang disusun oleh guru produktif mata pelajaran instalasi tenaga listrik di SMK Negeri 5 dan SMK Negeri 26 Jakarta belum pernah dianalisis, baik dari segi kualitas teknis maupun keterkaitannya dengan ketercapaian Tujuan Pembelajaran (TP) yang dirumuskan berdasarkan Capaian Pembelajaran (CP).
2. Kecenderungan guru menggunakan kembali soal-soal dari tahun sebelumnya tanpa melakukan analisis.
3. Soal yang digunakan selama ini belum tervalidasi, sehingga belum ada kepastian apakah butir soal telah memenuhi standar pengukuran yang baik.
4. Banyak soal ujian yang belum di uji validitas dan reabilitasnya, sudah digunakan oleh guru.
5. Beberapa siswa kesulitan menjawab soal karena tingkat kesukarannya tidak sesuai dengan kemampuan mereka. Ketidakseimbangan antara tingkat kesukaran soal dan kemampuan siswa menyebabkan hasil evaluasi tidak mencerminkan kompetensi secara objektif dan adil.

1.3. Pembatasan Masalah

Berdasarkan latar belakang masalah dan identifikasi masalah yang ada, maka batasan masalah dapat dibatasi pada masalah-masalah yang terkait dengan pembuatan soal. Sehingga landasan masalah dapat diuraikan sebagai berikut:

1. Dalam penelitian ini, peneliti membatasi identifikasi masalah nomor 4 yaitu pengujian validitas, reliabilitas, daya beda, tingkat kesukaran, serta efektivitas pengecoh soal buatan guru.
2. Mengingat banyaknya metode analisis untuk pengujian validitas dan reliabilitas maka, dalam penelitian ini akan peneliti batasi pada uji validitas dengan analisis butir soal menggunakan pendekatan model Rasch, serta ditinjau sejauh mana soal yang dianalisis mendukung ketercapaian Tujuan Pembelajaran (TP) yang dirumuskan berdasarkan Capaian Pembelajaran (CP).

1.4. Rumusan Masalah

Berdasarkan penjabaran dalam identifikasi masalah, maka dirumuskan masalah penelitian sebagai berikut:

1. Bagaimana kualitas soal Ujian Akhir Semester (UAS) buatan guru untuk mata pelajaran Instalasi Tenaga Listrik kelas XI di SMK Negeri Jakarta ditinjau dari validitas, reliabilitas, daya beda, tingkat kesukaran, dan efektivitas pengecoh menggunakan model Rasch?
2. Sejauh mana soal tersebut mendukung ketercapaian Tujuan Pembelajaran (TP) yang merujuk pada Capaian Pembelajaran (CP) dalam Kurikulum Merdeka?

1.5. Tujuan Penelitian

Berdasarkan permasalahan yang telah ditetapkan sebelumnya, maka dapat ditentukan tujuan dari penelitian ini, yaitu:

1. Mengetahui kualitas butir soal Ujian Akhir Semester (UAS) buatan guru untuk mata pelajaran Instalasi Tenaga Listrik kelas XI di SMK Negeri Jakarta ditinjau dari validitas, reliabilitas, daya beda, tingkat kesukaran, dan efektivitas pengecoh menggunakan model Rasch.
2. Mengevaluasi sejauh mana soal-soal tersebut dapat mendukung pencapaian Tujuan Pembelajaran (TP) dan Capaian Pembelajaran (CP) sesuai dengan Kurikulum Merdeka.

1.6. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut :

1. Secara Teoritis

Penelitian ini bertujuan untuk memperdalam pemahaman dalam bidang pendidikan, terutama dalam menganalisis kualitas soal evaluasi dengan pendekatan Teori Respon Butir (*IRT*). Penelitian ini mencakup validitas, reliabilitas, tingkat kesukaran, daya pembeda, serta efektivitas pengecoh dalam instrumen evaluasi. Selain itu, penelitian ini memberikan kontribusi dalam menilai sejauh mana soal evaluasi mendukung pencapaian Tujuan Pembelajaran (TP) dan Capaian Pembelajaran (CP). Hasil penelitian ini diharapkan menjadi dasar bagi penelitian lebih lanjut dalam pengembangan evaluasi pembelajaran yang lebih objektif dan akurat.

2. Secara Praktis

a. Bagi Peneliti

Sebagai implementasi dari pengetahuan dan keterampilan yang telah dipelajari selama berkuliah di Pendidikan Teknik Elektro Universitas Negeri Jakarta, serta sebagai syarat memperoleh gelar sarjana dalam rangka menyelesaikan studi di Universitas Negeri Jakarta. Selain itu, penelitian ini dapat menjadi pengalaman dalam menerapkan metode analisis butir soal berbasis *IRT* untuk mengevaluasi kualitas soal secara objektif.

b. Bagi Guru

Penelitian ini membantu guru dalam mengelola bank soal dengan lebih baik. Soal-soal yang telah dianalisis dapat digunakan kembali jika memenuhi kriteria kualitas, sedangkan soal yang kurang baik dapat direvisi. Dengan demikian, guru memiliki referensi soal yang terstandar untuk digunakan dalam asesmen berikutnya.

c. Bagi Sekolah

Hasil penelitian menjadi dasar bagi sekolah untuk menerapkan analisis butir soal secara rutin guna meningkatkan kualitas asesmen dan mendukung pembelajaran yang lebih tepat sasaran di pendidikan kejuruan.