

**ANALISIS TINGKAT KESULITAN LEVEL KOGNITIF PADA SOAL PENILAIAN
SUMATIF BAHASA MANDARIN KELAS XI SMA SEMEN GRESIK TAHUN
AJARAN 2023/2024**

对SEMEN GRESIK学校高中十一年级2023-2024学年汉语试卷

**ANALYSIS OF COGNITIVE LEVEL DIFFICULTY IN SUMMATIVE
ASSESSMENT QUESTIONS FOR MANDARIN LANGUAGE IN CLASS XI AT
SEMEN GRESIK HIGH SCHOOL FOR THE ACADEMIC YEAR 2023/2024**

Qonita Lukiaffa Azmi¹

Universitas Negeri Surabaya
qonita.20043@mhs.unesa.ac.id

Urip Zaenal Fanani²

Universitas Negeri Surabaya
zaenalfanani@unesa.ac.id

Abstrak

Analisis butir soal dilakukan untuk mengetahui kualitas setiap butir soal dari keseluruhan soal. Analisis butir soal ditinjau melalui validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh soal. Analisis butir soal menggunakan bantuan aplikasi Anates V4 dan *Microsoft Excel*. Penelitian ini bertujuan mendeskripsikan tingkat kesulitan level kognitif dan kualitas butir soal pada soal penilaian sumatif Bahasa Mandarin kelas XI SMA Semen Gresik. Penelitian ini menggunakan pendekatan penelitian deskriptif kuantitatif. Subjek penelitian ini adalah soal penilaian sumatif mata pelajaran Bahasa Mandari kelas XI SMA Semen Gresik. Pengumpulan data dengan menggunakan teknik dokumentasi. Berdasarkan hasil penelitian menyimpulkan bahwa tingkat kesulitan level kognitif pada soal penilaian sumatif memiliki tingkat kognitif C2 sebanyak 22,22% dengan jumlah 10 butir soal, C3 11,11% dengan jumlah 20 butir soal, dan C4 11,11% dengan jumlah 5 butir. Soal yang memiliki validitas tinggi 47%, soal yang memiliki validitas sangat tinggi 40%, dan soal yang memiliki validitas rendah 20%. Reliabilitas pada keseluruhan butir soal memiliki tingkat reliabilitas yang baik. Tingkat kesukaran soal dengan kesulitan sangat mudah 40%, soal kategori mudah 40%, dan soal kategori sedang 20%. Daya pembeda terdapat 67% soal dengan daya pembeda baik, dan 17% soal dengan daya pembeda cukup. Efektivitas pengecoh soal pilihan ganda terdapat dua butir soal yang memiliki kualitas pengecoh buruk atau tidak berfungsi.

Kata Kunci: Analisis butir soal, penilaian sumatif, Anates V4

摘要

进行项目分析，以评估一组问题中每个项目的质量。根据分析的有效性、可靠性、判别力、难度水平和问题中干扰物的有效性对分析进行评价。项目分析使用 Anates V4 应用程序和微软 Excel 进行。本研究旨在分析和描述 Semen Gresik 高中十一年级普通话终结性评估试题的认知困难程度和试题质量。本研究采用定量描述的方法，以 Semen Gresik 高中十一年级普通话期末考题为研究对象。数据收集是利用访谈技术和文件技术进行的。

根据研究结果得出结论，总结性评估问题中的认知困难程度如下：C2 级别为 22.22%，共有 10 项；C3 级别为 11.11%，共有 20 项；C4 级别为 11.41%，共有 5 项。有效性高的问题占 47%，有效性极高的问题占 40%，有效性低的问题占 20%。所有项目的可靠性表现出良好的可靠性水平。题目的难度等级分为非常容易（40%）、容易（40%）和中等（20%）。歧视力分析显示，67% 的问题具有良好的歧视力，17% 具有充分的歧视力；16% 具有较差的歧视能力。多项选择题中干扰物的有效性表明，有两项干扰物很差或不起作用。

关键词：项目分析，总结性评估，Anates V4

Abstract

Item analysis is conducted to assess the quality of each item in a set of questions. The analysis is evaluated based on validity, reliability, discriminative power, difficulty level, and the effectiveness of distractors in the questions. Item analysis is performed using the Anates V4 application and Microsoft Excel. This study aims to analyze and describe the cognitive difficulty levels and the quality of the items in the summative assessment questions for Mandarin Language in class XI at SMA Semen Gresik. The research employs a quantitative descriptive approach, with the subject being the summative assessment questions for Mandarin Language in class XI at SMA Semen Gresik. Data collection was conducted using interview techniques and documentation techniques. Based on the research results, it was concluded that the cognitive difficulty levels in the summative assessment questions are as follows: C2 level at 22.22% with 10 items, C3 level at 11.11% with 20 items, and C4 level at 11.11% with 5 items. Questions with high validity account for 47%, questions with very high validity make up 40%, and questions with low validity comprise 20%. The reliability of all items shows a good level of reliability. The difficulty level of the questions is categorized as very easy at 40%, easy at 40%, and moderate at 20%. The discriminative power shows that 67% of the questions have good discriminative power, 17% have adequate discriminative power, and 16% have poor discriminative power. The effectiveness of distractors in multiple-choice questions reveals that two items have poor or non-functioning distractors.

Keywords: Item analysis, summative assessment, Anates V4

1. PENDAHULUAN

1.1 Latar Belakang

Bahasa adalah sebuah sarana untuk menyerap dan mengembangkan pengetahuan serta merupakan alat komunikasi yang efektif untuk

melakukan interaksi dengan lawan bicara disebut bahasa. Ada banyak sekali keragaman bahasa di dunia. Dunia pendidikan telah mengembangkan berbagai macam bahasa asing sebagai dampak dari era globalisasi yang sangat mempengaruhi

satu bahasa terhadap bahasa yang lain.

Salah satu bahasa asing yang semakin populer dan penting dalam pengajaran bahasa adalah bahasa mandarin. Dalam mempelajari bahasa Mandarin peserta didik perlu mengenal dan memiliki keterampilan berbahasa Mandarin yang mencakup 听 *tīng* (menyimak), 说 *shuō* (berbicara), 读 *dú* (membaca), serta 写 *xiě* (menulis). Pada pembelajaran Bahasa Mandarin, peserta didik diwajibkan memiliki kemampuan berbahasa yang meliputi 4 keterampilan berbahasa tersebut (Fuo Ruomei & Agustian, 2017:24). Untuk mencapai kemampuan keterampilan berbahasa Mandarin perlunya dilakukan penilaian, agar dapat mengetahui kemampuan dalam berbahasa pada peserta didik.

Sekolah melakukan penilaian terhadap peserta didik melalui tes atau *assessment*. Diadakannya penilaian merupakan sebuah bentuk yang dari proses pembelajaran, serta sebagai gambaran usaha dari peserta didik terhadap apa yang telah dicapai dalam proses pembelajara.

Evaluasi hasil belajar dilakukan pada setiap mata pelajaran, pertengahan, dan di akhir pembelajaran, fungsinya untuk mengukur *skill* dan memahami materi yang telah diberikan dan diterima oleh peserta didik. Menurut Sugiyono, (2013:52) tes merupakan suatu metode pengukuran dan evaluasi dalam dunia pendidikan melalui pemberian tugas atau lembar kerja pada peserta didik. Oleh sebab itu, analisis butir soal penting untuk dilakukan terlebih dahulu supaya dapat mengetahui kualitas dari butir soal yang akan dibuat.

Analisis butir soal ialah kegiatan dibutuhkan oleh pendidik guna memberikan peningkatan terhadap

kualitas butir soal yang sudah dirancang. Adapun penyajian soal haruslah soal- soal yang bisa merangsang tingkat berpikir peserta didik. Soal-soal tersebut juga wajib mencakup isi materi/topik yang sudah diajarkan sebelumnya.

Soal yang berkualitas adalah soal yang mampu memberi informasi terkait materi atau topik yang telah diajarkan sebelumnya pada pembaca dengan sangat tepat. Kualitas soal baik dari keseluruhan ataupun butir soal bisa diamati dengan melakukan analisis kualitas (Arifin, 2017:246). Dilakukannya analisis butir soal berguna dalam mengidentifikasi butir soal yang bisa dipakai kembali juga soal yang perlu diperbaiki ataupun diganti dengan cara menghitung beberapa aspek, misalnya tingkat kesukaran, reliabilitas, efektivitas pengecoh, daya pembeda, serta validasi (Arifin, 2017:246-280).

Penilaian dilaksanakan guna mengetahui pengetahuan masing-masing peserta didik dalam penugasan dan pemahaman dari topik pembelajaran yang telah diajarkan selama proses pembelajaran disebut penilaian kognitif. Dalam menganalisa tingkat kognitif soal, pendidik hendaknya melakukan pengujian dari kualitas butir soal. Suatu tes yang digunakan untuk patokan alat pengukur pencapaian bisa dinilai baik maka wajib memenuhi persyaratan tes, yakni harus mempunyai: objektivitas, ekonomis, praktikabilitas, reliabilitas, dan validitas (Arikunto, 2013:94). Validitas erat kaitannya dengan tepat atau tidaknya butir soal saat melakukan pengukuran terhadap sesuatu serta reliabilitas hubungannya ialah terdapat pada kepercayaan. Keahlian peserta didik ketika menjawab suatu pertanyaan dapat menentukan tingkat kesulitan dari

butir soal, bukan pendidikanya, sedangkan kemampuan butir-butir soal tersebut dalam membedakan antara peserta didik yang sudah menguasai isinya dan peserta didik yang belum menguasainya dikenal dengan daya beda.

Penilaian pada hasil belajar peserta didik yang telah diimplementasikan sejauh ini di SMA Semen Gresik secara menyeluruh belum mengamati kualitas butir tes. Adapun didasarkan pada hasil wawancara singkat dengan pendidik mata pelajaran bahasa Mandarin di SMA Semen Gresik, maka didapatkan informasi bahwa pendidik melakukan analisis butir soal namun tidak terfokus pada tingkat reliabilitas, daya pembeda, validitas, efektivitas pengecoh soal, dan tingkat kesulitan. Bersumber pada hasil wawancara tersebut, perlu dilakukannya analisis butir soal yang fokusnya pada hal tersebut. Butir soal yang baik bisa dipakai untuk keperluan tes selanjutnya serta dapat disisipkan pada bank soal. Apabila butir soal kurang baik atau belum baik, revisi atau pembenahan ulang dapat dilakukan, serta dipakai kembali untuk tes selanjutnya, namun jika butir soal tidak berkualitas ataupun tidak baik sebaiknya tidak dipakai untuk tes berikutnya.

Berdasarkan pembahasan pada latar belakang tersebut, peneliti tertarik meneliti analisis butir soal sumatif Bahasa Mandarin yang terdapat di SMA Semen Gresik. Peneliti memilih soal tes sumatif karena bentuk soal yang lebih bervariasi, terdapat pilihan ganda biasa, pilihan ganda kompleks, mencocokkan, benar salah, dan menyusun kalimat. Pemilihan soal tes sumatif kelas XI SMA Semen Gresik tahun ajaran 2023/2024 sebagai objek

penelitian adalah penilaian sumatif ini menggunakan soal yang lebih bervariasi, sedangkan soal formatif soal yang disajikan kurang bervariasi, dan terlalu sedikit. Sebab itu, soal formatif tidak digunakan menjadi objek penelitian. Analisis dilakukan terhadap aspek daya pembeda, reliabilitas, tingkat kesukaran, validitas, juga efektivitas pengecoh soal atau *distractor* (Widiyanto, 2018:43).

1.2 Rumusan Masalah

Adapun rumusan masalah yang ditulis oleh peneliti yang didasarkan pada latar belakang yang telah dijabarkan, diantaranya sebagai berikut:

1. Bagaimana tingkat kesulitan level kognitif pada soal penilaian sumatif Bahasa Mandarin kelas XI di SMA Semen Gresik Tahun Ajaran 2023/2024?
2. Bagaimana validitas soal penilaian sumatif Bahasa Mandarin kelas XI di SMA Semen Gresik Tahun Ajaran 2023/2024?
3. Bagaimana reliabilitas soal penilaian sumatif Bahasa Mandarin kelas XI di SMA Semen Gresik Tahun Ajaran 2023/2024?
4. Bagaimana daya pembeda soal penilaian sumatif Bahasa Mandarin kelas XI di SMA Semen Gresik Tahun Ajaran 2023/2024?
5. Bagaimana efektivitas pengecoh soal penilaian sumatif Bahasa Mandarin kelas XI di SMA Semen Gresik Tahun Ajaran 2023/2024?

1.3 Tujuan Penelitian

Adapun tujuan penelitian yang diangkat oleh peneliti yakni sebagai berikut:

1. Mendeskripsikan tingkat kesulitan level kognitif pada soal penilaian sumatif Bahasa Mandarin kelas XI

di SMA Semen Gresik Tahun Ajaran 2023/2024.

2. Mendeskripsikan validitas soal penilaian sumatif Bahasa Mandarin kelas XI di SMA Semen Gresik Tahun Ajaran 2023/2024
3. Mendeskripsikan reliabilitas soal penilaian sumatif Bahasa Mandarin kelas XI di SMA Semen Gresik Tahun Ajaran 2023/2024
4. Mendeskripsikan daya pembeda soal penilaian sumatif Bahasa Mandarin kelas XI di SMA Semen Gresik Tahun Ajaran 2023/2024
5. Mendeskripsikan efektivitas pengecoh pada soal penilaian sumatif Bahasa Mandarin kelas XI di SMA Semen Gresik Tahun Ajaran 2023/2024.

1.4 Teori

Suatu proses yang saling berkesinambungan dengan hasil belajar peserta didik supaya dapat mengumpulkan informasi maupun data mengenai proses dan hasil pembelajaran, penilaian bersifat kualitatif. Penilaian didasarkan pada kriteria pertimbangan tertentu yang dituangkan dalam bentuk keputusan.

Berdasarkan penilaian yang telah dijabarkan di atas, maka definisi penilaian yakni suatu proses pengambilan keputusan berdasarkan data yang dikumpulkan dari mengukur hasil belajar peserta didik dengan menggunakan instrumen non tes dan tes. Umumnya, penilaian dikelompokkan dalam 2 macam yakni penilaian formatif dan sumatif.

1.4.1 Penilaian Sumatif

Penilaian yang dilakukan ketika seluruh materi pelajaran telah dianggap telah dipelajari,

atau ketika suatu unit pengalaman pembelajaran telah diselesaikan disebut penilaian sumatif. Tujuan dari penilaian tersebut yakni menentukan nilai yang didasarkan pada tingkatan hasil belajar peserta didik (Arifin, 2017:36). Cakupan penilaian sumatif melebihi pokok bahasan yang bertujuan untuk mengetahui peserta didik mampu menyesuaikan dari suatu unit pembelajaran ke unit berikutnya.

Orientasi penilaian yakni pada produk, tidak pada prosesnya (Sudjana, 2011:5). Penggunaan tes sering digunakan pada akhir suatu periode pembelajaran pada evaluasi sumatif. Tes tersebut meliputi sebagian atau semua materi yang sudah diberikan pada peserta didik dalam satu semester.

Adapun pendapat Arikunto (2013:53) tes sumatif ialah suatu tes yang pelaksanaannya dilakukan setelah pembelajaran ataupun di akhir semester. Adapun manfaat dari tes sumatif yaitu, yang pertama adalah sebagai penentu nilai, yang kedua sebagai penentu seseorang peserta didik bisa atau tidak ketika mengikuti kelompok pada program selanjutnya, dan yang ketiga adalah sebagai pengisi catatan kemajuan belajar peserta didik.

1.4.2 Fungsi Penilaian

Menurut Tim Direktorat Pembinaan SMP (2017:9-10) fungsi penilaian yaitu meningkatkan kemampuan peserta didik dalam proses belajar dan mengajar, pelaksanaan penilaian menggunakan 3 pendekatan,

yakni:

1. *Assesment of learning*.
Assesment of learning (penilaian akhir pembelajaran) ialah penilaian yang pelaksanaannya sesudah proses pembelajaran usai.
2. *Assesment for learning*.
Assesment for learning (penilaian untuk pembelajaran) pelaksanaannya selama proses pembelajaran serta biasanya digunakan sebagai landasan dalam meningkatkan proses pendidikan. Pada *assesment for learning* pendidik menilai proses belajar peserta didik, mencatat kemajuan mereka, dan mengevaluasi pembelajaran mereka.
3. *Assesment as learning*.
Assesment as learning (penilaian sebagai pembelajaran) mempunyai fungsi yang hampir sama dengan *assesment for learning*, fungsinya sebagai alat formatif yang diterapkan sepanjang proses pembelajaran dan ditentukan oleh hasil penilaian. Perbedaan diantara keduanya yakni, *assesment as learning* yang mencakup peserta didik secara aktif pada kegiatan penilaian tersebut.

Menurut Sudjana (2011:3-4) penilaian yang dilihat dari berbagai macam segi dalam sistem pendidikan, sehingga terdapat beberapa fungsi dari penilaian diantaranya sebagai berikut:

1. Penilaian berfungsi sebagai sarana untuk memastikan tercapai atau tidaknya tujuan

pembelajaran.

2. Evaluasi berfungsi sebagai masukan untuk perbaikan proses belajar mengajar.
3. Penilaian juga berfungsi sebagai landasan selama penyusunan laporan kemajuan belajar peserta didik.

1.4.3 Tes

Tes adalah suatu proses yang menggunakan pedoman dan prosedur yang telah ditetapkan untuk mengukur atau menemukan sesuatu dalam suatu suasana (Arikunto, 2013:67). Tes adalah instrumen atau prosedur yang digunakan untuk melakukan pengukuran yang terdiri dari serangkaian kegiatan, pernyataan, atau pertanyaan yang wajib diselesaikan peserta didik supaya dapat mengukur berbagai elemen kinerja mereka. (Arifin, 2017:118).

Menurut Magdalena dkk (2021:278-280) tes ialah sebuah alat yang efektif untuk mengukur tujuan pembelajaran telah tercapai melalui penilaian. Adapun tes dikelompokkan menjadi 2 jenis yakni, objektif dan tes subjektif dengan penjelasan sebagai berikut:

1. Tes Objektif

Tes objektif adalah tes yang dirancang sedemikian rupa sehingga hasil tes dapat dievaluasi secara tidak memihak. Pada tes objektif juga dibagi menjadi beberapa kelompok yakni, pilihan ganda, melengkapi, mengisi titik-titik, benar salah, dan menjodohkan.

2. Tes Subjektif

Tes subjektif sering disebut dengan tes uraian

karena memberikan kesempatan kepada peserta didik untuk memilih dan menetapkan jawabannya sendiri, sehingga menimbulkan data jawaban peserta didik yang beragam dan subjektivitas dalam proses evaluasi (Magdalena dkk 2021:278-280).

1.4.4 Analisis Butir Soal

Analisis butir soal adalah proses pemeriksaan soal hingga menghasilkan himpunan dengan keseimbangan kualitas yang sesuai, terdiri dari soal sederhana, sedang, dan sukar (sulit) menurut pernyataan dari (Sudjana 2011:134). Pendidik perlu melakukan analisis soal guna meningkatkan kualitas soal tes yang telah dikembangkan. Tujuan dari kegiatan ini adalah untuk mengumpulkan dan mengevaluasi tanggapan peserta didik untuk menentukan hasil dari setiap evaluasi (Ratnawulan & Rusdiana, 2014:168). Tujuan analisis item adalah untuk mengklasifikasikan pertanyaan menjadi baik, kurang baik, ataupun tidak baik.

Analisis butir soal dilakukan supaya dapat mengetahui mana soal yang baik, kurang baik, serta soal yang tidak baik. Dalam proses belajar mengajar, penganalisisan butir soal juga berperan karena pendidik menjadi lebih tahu mengenai kualitas soal yang akan dibuat bahkan soal yang telah dibuat.

1.4.5 Validitas

Menurut Arikunto (2013:79-84) akurat atau tidaknya hasil evaluasi terletak pada sebuah instrumen yang

digunakan sebagai alat ukur. Agar dapat mencapai hal tersebut, dibutuhkan alat ukur yang berkualitas dan baik, diantaranya seperti berikut :

1. Validitas Logis

Instrumen evaluasi yang sesuai kriteria validitas berdasarkan kesimpulan penalaran dikatakan mempunyai validitas logis. Kondisi valid dapat dinyatakan terpenuhi karena instrumen yang berkaitan telah disusun dengan baik, sesuai dengan ketentuan yang ada dan mengikuti teori. Suatu instrumen harus mencapai validitas isi dan validitas konstruk (*construct validity*), yang merupakan dua jenis validitas logis. Keadaan suatu instrumen yang dibuat tergantung pada isi pokok bahasan yang dinilai disebut dengan validitas isi. Validitas kontrak (*construct validity*) merupakan keadaan suatu instrumen yang dikembangkan dengan menggunakan faktor psikologis yang perlu dinilai.

2. Validitas Empiris

Instrumen yang telah diuji dari pengalaman, maka dapat dikatakan mempunyai validitas empiris. Adapun pembagian validitas empiris dikelompokkan dalam bentuk 2 validitas, yakni validitas “ada sekarang” (*concurrent validity*) dan validitas prediksi (*predictive validity*). Validitas “ada sekarang” (*concurrent validity*) dikatakan memiliki validitas empiris apabila hasilnya sesuai dengan pengalaman.

1.4.6 Reliabilitas

Reliabilitas adalah sejauh mana atau banyaknya derajat konsistensi suatu instrumen. Suatu reliabilitas yang kaitannya dengan pertanyaan, apakah suatu tes komprehensif serta bisa diandalkan dengan menyesuaikan standar yang telah ditentukan. Suatu tes bisa dinilai reliabel ketika kelompok yang sama diuji beberapa kali dan hasilnya tetap sama di waktu yang berbeda (Arifin, 2016:258). Kerlinger (1976) dalam Arifin (2016:258) menyatakan tiga kriteria yang dapat digunakan untuk mengukur reliabilitas: (*predictability*) menunjukkan kemampuan tes untuk meramalkan hasil, (*dependability*) menunjukkan tingkat keandalan atau kepastian tes, dan (*stability*) menunjukkan konsistensi tes dalam mengukur hal yang sama pada waktu yang berbeda. pada pengukuran selanjutnya.

1.4.7 Daya Pembeda

Kemampuan suatu soal untuk membedakan antara peserta didik yang kurang cerdas (yaitu belum menguasai isi) dan peserta didik yang lebih intelektual (yaitu menguasai materi) dikenal dengan istilah daya pembeda. Suatu proporsi dapat digunakan untuk menyatakan indeks kekuatan pembeda, semakin besar proporsinya, semakin baik pertanyaannya (Arifin, 2016:133).

Menurut Sudjana (2011:141), daya pembeda dapat dimanfaatkan pada saat memeriksa butir-butir soal untuk mengetahui apakah dapat

membedakan peserta didik pada kategori menguasai dan tidak menguasai. Jika suatu tes diberi pada peserta didik yang berprestasi tinggi namun mendapat nilai rendah, namun jika tes tersebut dibagikan pada peserta didik dalam hal prestasinya kurang maka hasilnya lebih tinggi, atau jika tes diberikan kepada kedua kategori peserta didik tersebut hasilnya sama saja, maka tes tersebut dianggap tidak memiliki daya pembeda.

1.4.8 Efektivitas Pengceoh

Distractor bisa diasumsikan baik atau efektif apabila *distractor* tersebut telah ditentukan paling sedikit 5% daripada jumlah keseluruhan peserta tes. Apabila soal berupa pilihan ganda terdapat alternatif pada jawaban sebagai pengecoh pada jawaban, sesuai dengan pernyataan Arifin (2016:279). Jika butir soal baik, maka pengecohnya akan dipilih dengan merata oleh peserta didik yang memberikan jawaban salah. Sebaliknya, jika butir soal kurang baik, maka pengecoh yang dipilih tidak merata oleh peserta didik. Oleh karena itu, jika didasarkan pada teori yang telah dijabarkan, kesimpulannya yakni efektivitas pengecoh bisa dikatakan baik jika pengecoh itu berfungsi dan telah dipilih minimal 5%.

1.4.9 Tingkat Kesulitan

Tingkat kesulitan atau tingkat kesukaran bertujuan untuk menunjukkan seberapa mudah atau sulitnya bentuk inkuiri tersebut. (Arikunto, 2013:207). Tingkat kesulitan suatu soal merupakan indikator

numerik seberapa mudah atau rumitnya soal tersebut. Menurut Sudjana (2011:135) tingkat kesukaran soal dapat diketahui dari kemampuan ataupun kapasitas peserta didik saat menanggapi suatu pertanyaan yang dapat digunakan untuk menentukan tingkat kesulitan suatu pertanyaan, bukan sudut pandang dari pendidik yang membuat soal. Aspek penting dari analisis kesukaran soal adalah mencari tahu berapa banyak soal yang mudah, berapa banyak yang sedang, dan berapa banyak yang menantang. Soal yang baik ialah soal dengan tingkat kesulitan atau kemudahannya menggunakan proporsi yang tepat. Peserta didik tidak dapat didorong untuk meningkatkan upaya pemecahan masalah mereka dengan mengerjakan soal yang terlalu mudah. Namun berkebalikan dengan soal yang terlalu sulit, hal tersebut akan menurunkan motivasi peserta didik dan membuat mereka enggan untuk mencoba lagi karena diluar kemampuan yang mereka miliki.

1.4.10 Taksonomi Bloom

Taksonomi secara etimologis dapat dipahami sebagai hukum untuk mengatur sesuatu; itu dibagi menjadi beberapa tingkatan yang mewakili tingkat kompleksitas. Tingkat kesulitan ini juga merupakan tantangan yang dihadapi selama proses belajar mengajar. Secara umum, Bloom dan rekan – rekannya mengembangkan tiga tingkatan tujuan pendidikan:

1. Perluasan kategori menjadi sederet tujuan.

2. Kelompok tingkah laku yang masih verbal.
3. Perilaku konkrit yang terdiri dari tes dan soal tes yang terdiri atas tugas-tugas yang bersangkutan.

Menurut Arifin (2014:130) terdapat 3 domain atau ranah yang susunannya pada urutan ke-2 berikutnya disebut taksonomi yakni:

1. *Cognitive domain* yakni ranah kognitif.
2. *Affective domain* yakni ranah afektif.
3. *Psychomotor domain* yakni ranah psikomotor.

1.4.11 Kognitif

Aktivitas kognitif mencakup hal-hal seperti memahami suatu peristiwa dan kemudian melanjutkan untuk memahami dan memahaminya. Hal tersebut berkaitan dengan proses belajar mengajar. Definisi lain dari fungsi kognitif ialah proses mental yang memungkinkan seseorang memahami, menilai, dan merefleksikan suatu peristiwa. Akibatnya, kemampuan kognitif dan kecerdasan saling terkait erat.

Kemampuan memperoleh makna dari item atau materi yang dipelajari dapat dipahami sebagai gagasan kognitif (Anggraini, 2012). Hal tersebut memperjelas sejauh mana peserta didik dapat menerima, mengasimilasi, serta memahami informasi yang diberikan pendidik pada mereka, atau seberapa jauh mereka bisa mempelajari dan memahami dari yang mereka baca, lihat, rasakan, atau alami sebagai akibat langsung dari tindakan mereka.

Semua proses mental seseorang yang terlibat dalam pendidikan dan pembelajaran yang memungkinkan mereka merefleksikan dan memahami suatu peristiwa disebut kognitif. Mengingat sebagian besar tugas belajar selalu mencakup mengingat dan berpikir, unsur kognitif sangat penting untuk pembelajaran yang efektif. Akibatnya, kemampuan kognitif dan tingkat kecerdasan sangat terkait.

2. METODE PENELITIAN

Jenis penelitian ini menggunakan penelitian kuantitatif untuk mengumpulkan data statistik atau angka yang dihitung melalui persentase kognitif, validitas, efektivitas pengecoh, daya pembeda, dan reliabilitas, daya pembeda. Analisis deskriptif merupakan pendekatan yang diimplementasikan pada penelitian ini. Penggunaan metode analisis deskriptif untuk mendeskripsikan dan menjelaskan apa yang terjadi. Adapun letak analisis deskriptif dalam penelitian ini bertujuan untuk mendeskripsikan tingkat kesulitan level kognitif, validitas, daya pembeda, reliabilitas, dan efektivitas pengecoh pada soal penilaian sumatif bahasa Mandarin kelas XI SMA Semen Gresik tahun ajaran 2023/2024.

Sumber data utama peneliti adalah lembar soal tes penilaian sumatif yang dibuat oleh pendidik mata pelajaran Bahasa Mandarin di SMA Semen Gresik dan jawaban hasil dari tes penilaian sumatif yang dikerjakan oleh peserta didik Kelas XI SMA Semen Gresik. Pada penelitian ini, hasil data penelitian berupa 2

bentuk data sebagai berikut:

1. Data berupa soal tes penilaian sumatif yang dibuat oleh pendidik mata pelajaran Bahasa Mandarin di SMA Semen Gresik.
2. Data berupa hasil jawaban dari soal tes penilaian sumatif yang dilakukan setelah mengisi soal tes selama proses kegiatan belajar mengajar oleh pendidik mata pelajaran Bahasa Mandarin di SMA Semen Gresik.

Peneliti menggunakan teknik metode dokumentasi dalam penelitian ini. Metode dokumentasi yakni menelusuri dokumen-dokumen seperti buku, catatan, dan lain- lain yang tujuannya untuk mendapatkan informasi mengenai objek atau variabel (Arikunto (2013), hal. 274). Metode dokumentasi yang terdapat pada penelitian ini adalah lembar soal tes penilaian dan kunci jawaban. Selain metode penelitian, instrumen penelitian merupakan hal yang penting. Pengukuran fenomena yang diteliti menggunakan suatu alat yang disebut instrumen penelitian. Adapun penggunaan instrumen oleh peneliti dalam penelitian ini adalah lembar soal tes penilaian dan kunci jawaban dan lembar hasil jawaban peserta didik.

Analisis data dalam penelitian ini diaplikasikan pada butir soal tes pada penilaian sumatif dalam pelajaran Bahasa Mandarin kelas XI di SMA Semen Gresik 2023/2024 dengan menggunakan teknik analisis secara deskriptif kuantitatif dan menggunakan bantuan aplikasi Anates V4 dan *Microsoft Excel*. Pencarian daya pembeda, validitas, dan reliabilitas menggunakan aplikasi Anates V4. Aplikasi Anates V4 digunakan untuk menghitung keseluruhan kriteria validitas, reliabilitas, daya pembeda, dan

efektivitas pengecoh.

3. HASIL DAN PEMBAHASAN

No	Kategori	No. Soal	Banyak Soal	Persentase
1.	Mengingat (C1)	Pilihan ganda : 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15		
		Pilihan ganda kompleks : 16, 17, 18, 19, 20		
		Menjodohkan : 21, 22, 23, 24, 25		
		Benar salah : 26a, 26b, 26c, 27a, 27b, 27c, 28a, 28b, 28c, 29a, 29b, 29c, 30a, 30b, 30c		
		uraian : 31, 32, 33, 34, 35		
2.	Memahami (C2)	Pilihan ganda : 1, 2, 3, 4, 5, 6, 7, 8, 9, 10	10	28,57%
3.	Mengaplikasikan (C3)	Pilihan ganda : 11, 12, 13, 14, 15	5	14,29%
		Pilihan ganda kompleks : 16, 17, 18, 19, 20	5	14,29%
		Menjodohkan : 21, 22, 23, 24, 25	5	14,29%
		Benar salah : 26a, 26b, 26c, 27a, 27b, 27c, 28a, 28b, 28c, 29a, 29b, 29c, 30a, 30b,	5	14,29%

		30c		
4.	Menganalisis (C4)	uraian : 31, 32, 33, 34, 35	5	14,29%
5.	Mengevaluasi (C5)	-	0	0,00
6.	Mencipta (C6)	-	0	0,00
Jumlah soal			35	100

Penelitian ini merupakan jenis penelitian kuantitatif dengan pendekatan analisis deskriptif. Penelitian dilakukan di SMA Semen Gresik dengan menggunakan hasil jawaban 2 kelas yaitu, kelas XI.1 dan kelas XI.5, agar mendapatkan hasil yang valid peneliti menggunakan 2 kelas dengan teknik acak atau *simple random sampling*. Pada tahap awal sebelum dilaksanakan penelitian, peneliti telah melakukan *review* terhadap beberapa soal penilaian sumatif, namun terdapat beberapa soal yang sangat sedikit jumlahnya, dan tidak beragam jenisnya atau tidak variatif. Peneliti memilih soal penilaian sumatif Bahasa Mandarin di SMA Gresik karena bentuk soal yang bervariasi.

Terdapat beberapa persamaan dan perbedaan pada penelitian ini dengan penelitian terdahulu yang relevan. Penelitian ini memiliki kesamaan yaitu peneliti sama-sama membahas analisis butir soal yang ditinjau melalui validitas, reliabilitas, daya pembeda dan efektivitas pengecoh. Pada metode penelitian peneliti juga menggunakan metode deskriptif kuantitatif, adapun perbedaan pada penelitian ini dan penelitian terdahulu yaitu tidak membahas mengenai tingkat kesulitan

butir soal level kognitif dengan menggunakan taksonomi Bloom, selain itu pada teknik perhitungan peneliti terdahulu menggunakan aplikasi IBM SPSS *Statistics* versi 25, dan penelitian ini menggunakan aplikasi Anates V4 dan *Microsoft Excel*.

Perhitungan dalam penelitian ini dilakukan dengan bantuan aplikasi Anates V4 dan *Microsoft Excel*. Aplikasi Anates V4 bertujuan untuk mendeskripsikan validitas, reliabilitas, daya pembeda, dan efektivitas pengecoh pada soal penilaian sumatif, dengan menggunakan bantuan aplikasi Anates V4 peneliti memasukkan data yang berupa hasil jawaban masing-masing milik peserta didik kelas XI.1 dan XI.5, kemudian aplikasi Anates V4 akan mengeluarkan hasil validitas, reliabilitas, daya pembeda, dan efektivitas pengecoh soal penilaian sumatif, setelah itu peneliti menghitung jumlah persentase menggunakan *Microsoft Excel* berdasarkan jumlah keseluruhan tingkat kognitif soal dengan menggunakan rumus yang telah disebutkan.

Berdasarkan tingkat kognitif, soal penilaian sumatif Bahasa Mandarin memiliki tingkat kognitif C2, C3, dan C4. Pada tingkatan kognitif memahami terdapat pada soal dengan proses kognitif menafsirkan, membandingkan, mengimplementasikan, dan membedakan. Tingkat kognitif yang mendominasi, yaitu menafsirkan, dikarenakan terdapat banyak soal yang menuntut peserta didik untuk

menerjemahkan kata bahasa Mandarin ke dalam bahasa Indonesia. Analisis tingkat kesulitan level kognitif yang mengacu pada Taksonomi Bloom pada soal penilaian sumatif memiliki tingkat kognitif C1 yaitu pada seluruh soal, C2 dengan persentase 28,57% jumlah 10 butir soal, C3 dengan persentase 66,67% sebanyak 20 butir soal, dan C4 dengan persentase 14,29% sebanyak 5 butir soal.

Dalam pembelajaran di tingkat SMA, penyusunan soal seharusnya lebih berorientasi pada level kognitif tinggi, khususnya kategori C4 (menganalisis), C5 (mengevaluasi), dan C6 (mencipta), guna mendorong peserta didik untuk berpikir kritis, logis, dan inovatif sesuai dengan prinsip Higher Order Thinking Skills (HOTS). Namun, dalam implementasinya, pada pelaksanaan penilaian sumatif Bahasa Mandarin kelas XI semester genap di SMA Semen Gresik tahun 2023/2024 masih banyak soal yang didominasi oleh kategori C1 (mengingat), C2 (memahami), dan C3 (mengaplikasikan), yang belum sepenuhnya mencerminkan standar HOTS. Kondisi ini dapat berdampak pada rendahnya kemampuan peserta didik dalam menganalisis permasalahan kompleks, mengevaluasi informasi secara mendalam, serta menghasilkan solusi yang inovatif. Oleh karena itu, diperlukan upaya peningkatan kualitas penyusunan soal dengan menitikberatkan pada keterampilan berpikir tingkat tinggi agar pembelajaran lebih bermakna dan

selaras dengan tuntutan kurikulum modern.

Analisis butir soal pilihan ganda pada soal penilaian sumatif mata pelajaran Bahasa Mandarin kelas XI di SMA Semen Gresik dapat diketahui dari jumlah 15 butir soal pilihan ganda, dimana sebanyak 6 butir soal yang tergolong rendah validitasnya, 4 butir soal tergolong tinggi validitasnya, serta 5 butir soal tergolong sangat tinggi validitasnya. Pada soal nomor 2, 4, 5, 6, 7, 8, 9, 11, 14, dan 15 memiliki validitas tinggi sehingga butir soal dikatakan valid dan dapat digunakan kembali, sedangkan butir soal yang mempunyai tingkat validitas rendah lebih baik direvisi atau diganti. Kategori reliabilitas pada soal pilihan ganda berdasarkan koefisien korelasi dapat dinyatakan bahwa pada soal tersebut mempunyai reliabilitas tinggi, soal penilaian sumatif Bahasa Mandarin memiliki koefisien reliabilitas 0,73 dan dapat dinyatakan bahwa soal tersebut mempunyai tingkat konsistensi yang tinggi. Tingkat kesukaran pada soal pilihan ganda mempunyai tingkat kesulitan yang sangat mudah dengan total 8 soal, sebanyak 3 butir soal digolongkan dalam kategori mudah, dan 4 butir soal yang tergolong sedang. Dari analisis tingkat kesukaran butir soal dengan penyebaran 53% butir soal sangat mudah, 27% butir soal sedang, serta 20% butir soal mudah. Tingkat kesulitan yang tidak merata atau tidak merata pada soal pilihan ganda mengakibatkan kualitas soal yang buruk. Daya pembeda pada soal soal pilihan ganda menyebutkan dua butir

soal pilihan ganda dikategorikan mempunyai daya pembeda sangat baik, lima butir soal mempunyai daya pembeda baik, tujuh butir mempunyai daya pembeda cukup, dan satu butir mempunyai daya pembeda yang buruk. Adapun daya pembeda berdasarkan soal-soal tersebut, maka soal nomor 6, 7, 8, 12, dan 15 mencakup 33% butir soal yang cukup, artinya soal-soal tersebut mampu membedakan kelompok atas dan kelompok bawah. Mengingat indeks daya pembeda butir soal ini antara 0,41 dan 1,00, maka pertanyaan ini dapat digolongkan mempunyai kualitas yang baik.

Analisis butir soal pilihan ganda kompleks pada soal penilaian sumatif mata pelajaran Bahasa Mandarin kelas XI di SMA Semen Gresik dapat diketahui dari 5 jumlah butir soal pilihan ganda kompleks terdapat 1 butir soal tergolong tinggi validitasnya, dan 4 butir soal tergolong sangat tinggi validitasnya. Pada soal nomor 2, 3, 4, dan 5 mempunyai validitas tinggi sehingga butir soal dikatakan valid dan dapat digunakan kembali, sedangkan butir soal yang mempunyai tingkat validitas rendah sebaiknya diperbaiki atau diganti. Kategori reliabilitas berdasarkan koefisien korelasi dapat dinyatakan bahwa pada soal pilihan ganda kompleks memiliki reliabilitas yang tinggi. Soal pilihan ganda kompleks pada penilaian sumatif Bahasa Mandarin memiliki koefisien reliabilitas 0,76 dan dapat dikatakan bahwa terdapat tingkat keseragaman yang tinggi. Tingkat kesukaran pada soal pilihan ganda kompleks terdapat satu butir soal tergolong sangat mudah,

dua soal tergolong mudah, dan dua soal tergolong sedang. Dari analisis tersebut menunjukkan bahwa terdapat 57% soal sedang, 29% soal mudah, dan 14% soal sangat mudah jika ditinjau dari tingkat kesulitannya. Daya pembeda pada soal pilihan ganda kompleks memiliki daya pembeda baik. Seluruh butir soal pada rangkaian soal ini mempunyai daya pembeda yang baik.

Analisis butir soal mencocokkan pada soal penilaian sumatif mata pelajaran Bahasa Mandarin kelas XI di SMA Semen Gresik dapat diketahui dari 5 jumlah butir soal mencocokkan seluruh butir soal tergolong sangat tinggi validitasnya. Pada soal nomor 1, 2, 3, 4, dan 5 memiliki validitas sangat tinggi sehingga butir soal dikatakan valid dan dapat digunakan kembali. Kategori reliabilitas berdasarkan koefisien korelasi dapat dinyatakan bahwa pada soal mencocokkan memiliki reliabilitas yang tinggi. Soal mencocokkan pada penilaian sumatif Bahasa Mandarin memiliki koefisien reliabilitas 0,72 dan dapat dikatakan bahwa terdapat tingkat keseragaman yang tinggi. Tingkat kesukaran pada soal mencocokkan terdapat satu soal yang dikategorikan sangat mudah, dua soal mudah, dan dua soal sedang. Dari analisis tersebut menunjukkan bahwa pada keseluruhan butir soal 100% tergolong sedang. Daya pembeda pada soal mencocokkan terdapat dua soal yang mempunyai daya pembeda baik dan tiga soal yang mempunyai daya pembeda sangat baik. Adapun hasil daya pembeda pada soal-soal tersebut, pada soal 3, 4, dan 5 mencakup 60% dari soal termasuk kategori yang baik.

Analisis butir soal benar salah pada soal penilaian sumatif mata pelajaran Bahasa Mandarin kelas XI di SMA Semen Gresik dapat diketahui dari 15 jumlah butir soal benar salah ada 11 butir soal yang tergolong rendah validitasnya, 2 butir soal tergolong tinggi validitasnya, dan 2 butir soal tergolong sangat tinggi validitasnya. soal nomor 5 dan 9 memiliki validitas tinggi, nomor 7 dan 8 memiliki validitas yang sangat tinggi. Kategori reliabilitas berdasarkan koefisien korelasi dapat dinyatakan bahwa pada soal benar salah memiliki reliabilitas yang tinggi. Soal benar salah pada penilaian sumatif Bahasa Mandarin memiliki koefisien reliabilitas 0,63 dan dapat dikatakan bahwa terdapat tingkat keseragaman yang tinggi. Tingkat kesukaran pada soal benar salah terdapat terdapat 9 butir soal yang tergolong sangat mudah dan 6 butir soal tergolong mudah. Dari analisis tersebut menunjukkan bahwa pada keseluruhan butir soal benar salah 60% butir soal sangat mudah dan 40% butir soal mudah. Daya pembeda pada soal benar salah mempunyai delapan butir soal yang berkategori mempunyai daya pembeda cukup, empat butir soal mempunyai daya pembeda tidak baik, dan tiga butir soal mempunyai daya baik. Daya pembeda dari soal tersebut didapatkan hasil butir soal 1, 2, 3, 4, 7, 9, 10, dan 14 merupakan 53% dari butir soal yang cukup.

Analisis butir soal uraian pada soal penilaian sumatif mata pelajaran Bahasa Mandarin kelas XI di SMA Semen Gresik dapat diketahui dari jumlah 4 butir soal uraian tergolong

memiliki validitas yang sangat tinggi dan 1 butir soal tergolong tinggi validitasnya. 1, 2, 3, dan 4 mempunyai validitas sangat tinggi, soal nomor 5 memiliki validitas tinggi. Kategori reliabilitas berdasarkan koefisien korelasi dapat dinyatakan bahwa pada soal uraian memiliki reliabilitas yang sangat tinggi. Soal uraian pada soal penilaian sumatif Bahasa Mandarin memiliki koefisien reliabilitas 0,82 dan dapat dikatakan bahwa terdapat tingkat keseragaman yang tinggi. Tingkat kesukaran pada soal uraian 5 butir soal yang tergolong mudah. Dari analisis tersebut menunjukkan bahwa pada keseluruhan butir uraian 100% tergolong butir soal mudah. Daya pembeda pada soal uraian mempunyai empat butir yang daya pembedanya baik dan satu butir yang daya pembedanya cukup, maka terdapat 80% % butir soal yang baik yaitu nomor 1, 2, 3, dan 4. Sedangkan 20% butir soal tergolong cukup.

Analisis validitas dan reliabilitas instrumen dalam penelitian ini menunjukkan adanya ketidakstabilan. Ketidakstabilan ini disebabkan oleh perbedaan karakteristik akademik antar kelas yakni, kelas XI 1 terdiri dari peserta didik dengan tingkat kemampuan berpikir yang umumnya tinggi, sementara kelas XI 5 memiliki tingkat kemampuan yang lebih beragam. Perbedaan karakteristik ini berpengaruh terhadap hasil analisis, terutama karena banyak butir soal yang tergolong mudah, sehingga validitas dan reliabilitas instrumen menjadi kurang konsisten di seluruh populasi yang diuji. Hasil analisis

reliabilitas dalam penelitian ini menunjukkan ketidakstabilan yang dapat dipengaruhi oleh faktor-faktor eksternal, salah satunya adalah karakteristik akademik peserta didik. Seperti perbedaan tingkat kemampuan berpikir di antara kelas yang diuji berpotensi menyebabkan variasi dalam respons terhadap butir soal. Hal ini berdampak pada ketidakstabilan reliabilitas yang diperoleh.

Ketidakstabilan validitas dan reliabilitas ini dapat disebabkan oleh beberapa faktor, antara lain variasi dalam sampel responden yang mencerminkan perbedaan tingkat pemahaman akademik. Variasi ini dapat mempengaruhi korelasi antara butir soal dengan total skor, sehingga mengurangi konsistensi hasil pengukuran. Selain itu, ukuran sampel yang kecil atau kurang representatif juga dapat menyebabkan hasil analisis yang tidak stabil atau fluktuatif. Oleh karena itu, dalam penelitian selanjutnya, diperlukan pemilihan sampel yang lebih heterogen serta instrumen yang memiliki tingkat kesulitan yang lebih bervariasi agar dapat meningkatkan validitas dan reliabilitas pengukuran.

Pada analisis pengecoh menunjukkan bahwa pengecoh pada soal yang dikategorikan baik cenderung dipilih secara merata oleh peserta didik yang memberikan jawaban salah. Sebaliknya, pada soal yang dikategorikan kurang baik, pengecoh tidak dipilih secara merata, sehingga fungsinya dalam mengarahkan peserta didik yang belum memahami materi menjadi

kurang efektif. Dari hasil analisis, ditemukan bahwa empat butir soal (40%) termasuk dalam kategori tidak baik, sepuluh butir soal (50%) termasuk dalam kategori kurang baik, dan hanya

satu butir soal (10%) yang memiliki pengecoh yang efektif. Efektivitas pengecoh dapat diukur dari sejauh mana peserta didik yang belum memahami konsep memilih alternatif jawaban pengecoh yang telah dirancang. Jika peserta didik cenderung memilih salah satu pengecoh yang disediakan, maka pengecoh tersebut bekerja secara efektif. Namun, ditemukan dua butir soal dengan pengecoh yang tidak berfungsi secara optimal, yakni soal nomor 1 dan 5. Pada kedua soal ini, terdapat pilihan jawaban yang sama sekali tidak dipilih oleh peserta didik, yang mengindikasikan bahwa pengecoh tidak berhasil menarik perhatian peserta didik yang kurang memahami materi. Hal ini menunjukkan bahwa butir soal tersebut perlu direvisi agar pengecoh dapat berfungsi dengan baik dalam membedakan pemahaman peserta didik.

4. SIMPULAN DAN SARAN

4.1 Simpulan

Berdasarkan hasil dan pembahasan berkaitan dengan analisis tingkat kesulitan level kognitif pada soal penilaian sumatif mata pelajaran Bahasa Mandarin kelas XI SMA Semen Gresik dapat disimpulkan pada poin-poin berikut:

1. Hasil analisis tingkat kesulitan level kognitif didasarkan pada

Taksonomi Bloom pada soal penilaian sumatif mempunyai tingkat kognitif C2 yakni 28,57% sebanyak 10 butir soal, C3 66,67% sebanyak 20 butir soal, serta C4 14,29% sebanyak 5 butir soal.

2. Validitas pada soal pilihan ganda terdapat 6 butir soal yang tergolong validitas rendah, 4 butir soal tergolong validitas tinggi, dan 5 butir soal tergolong sangat tinggi validitasnya. Validitas pada soal pilihan ganda kompleks sebanyak 1 butir soal tergolong validitas tinggi, juga 4 butir soal tergolong sangat tinggi validitasnya. Validitas pada soal mencocokkan tergolong validitas yang sangat tinggi sebanyak 5 butir soal. Validitas pada soal benar salah terdapat 15 butir soal tergolong validitas rendah, 2 butir soal tergolong validitas tinggi, serta 2 butir soal tergolong validitas sangat tinggi, dan validitas pada soal uraian terdapat 4 butir soal tergolong validitas sangat tinggi, dan 1 butir soal tergolong validitas tinggi.
3. Reliabilitas pada soal pilihan ganda dapat dipercaya atau reliabel karena mempunyai tingkat konsistensi yang tinggi yaitu 0,73. Reliabilitas yang terdapat pada soal pilihan ganda kompleks bisa dipercaya atau reliabel dikarenakan mempunyai tingkat konsistensi yang tinggi yakni 0,76. Reliabilitas pada soal pilihan mencocokkan dapat dipercaya atau reliabel karena mempunyai tingkat konsistensi yang tinggi yakni 0,72. Reliabilitas pada soal benar salah dapat dipercaya atau reliabel karena memiliki tingkat konsistensi tinggi yakni 0,78, dan Reliabilitas pada soal uraian dapat dipercaya atau reliabel karena mempunyai tingkat konsistensi yang tinggi yaitu 0,82.

4. Adapun sejumlah 80% butir soal pilihan ganda mempunyai daya pembeda yang cukup dan baik, maka daya pembeda tersebut dinilai baik. Daya pembeda yang terdapat di soal pilihan ganda kompleks tergolong baik karena butir soal 100% mempunyai daya pembeda yang baik. Daya pembeda pada soal mencocokkan tergolong baik sekali karena butir soal 60% mempunyai daya pembeda yang sangat baik. Daya pembeda pada soal benar salah tergolong cukup karena butir soal 53% mempunyai daya pembeda yang cukup. Pada soal uraian daya pembedanya tergolong baik karena 80% dari butir soal mempunyai daya pembeda tergolong baik. Efektivitas pengecoh dalam soal pilihan ganda terdapat 2 butir soal mempunyai pengecoh (distraktor) yang ada pada butir soal tersebut tidak berguna.

4.2 Saran

Agar penelitian dapat berkembang lebih baik, maka diperlukan saran. Adapun beberapa saran berdasarkan hasil analisis yang telah dilaksanakan oleh peneliti sebagai berikut. Pertama, perlu adanya peningkatan kualitas soal dengan memperhatikan tingkat kesukaran butir soal serta jenis soal yang lebih bervariasi. Pendidik perlu menambahkan berbagai macam soal dengan tingkat kualitas yang telah diuji secara valid. Kedua, Pada pilihan jawaban atau opsi sebagai pengecoh lebih diperhatikan lagi, supaya seluruh butir soal mempunyai pengecoh yang baik dan berfungsi. Ketiga, Analisis butir soal perlu dilakukan oleh pendidik

agar lebih tahu dan berkenan mengganti dan merevisi butir soal yang kurang baik.

DAFTAR PUSTAKA

- Abidin, S. F. F. (2021). *Analisis soal penilaian akhir semester (pas) bahasa mandarin dengan kriteria higher order thinking skill (hots)* (Doctoral dissertation, Universitas Negeri Malang).
- Abduh, Moch. (2019). *Panduan Penilaian Tes Tertulis*. Jakarta: Pusat Penilaian Pendidikan Ade, Hossiana, Inas, Naurah, Niken, Ade. (2021). *Penilaian Sumatif dan Penilaian Formatif Pembelajaran Online*. Ahmad Sabri, (2005). *Strategi Belajar Mengajar*. Ciputat Press.
- Aini, D. N., Afifah, Y. N., & Sakti, K. F. L. (2022). *Authentic Assessment pada Keterampilan Menulis Peserta didik Bahasa Mandarin*. Edukasi: Jurnal Pendidikan, 20(2), 300-314.
- Arikunto, Suharsimi. (2018). *Dasar-Dasar Evaluasi Pendidikan : edisi 3*. Jakarta: Rineka Cipta.
- Arikunto, Suharsimi. (2013). *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: PT Rineka Cipta
- Arifin, Zainal. (2017). *Evaluasi Pembelajaran*. Bandung: PT Remaja Rosdakarya
- Anderson, L. W. (1999). *Rethinking Bloom's Taxonomy: Implications for Testing and Assessment*.
- Anderson, W. L., & Krathwohl, R. D. (2010). *Kerangka Landasan Pembelajaran dan Asesmen Revisi Taksonomi Bloom Revisi*. Terjemahan oleh Prihantoro, A.
- Andrias, S. P., Saputra, A., Rais, R., Abdurrahman, S. K., Purnomo, A. C., Efrida Ita, S. S., ... & Fitriani, M. (2023). *MANAJEMEN PENDIDIKAN*.



- Selat Media.
- Aniswita, Neviyarni. (2020). *Perkembangan Kognitif, Bahasa, Perkembangan Sosio- emosional, dan Implikasinya Dalam Pembelajaran*. Djaali. (2020) *Metodologi Penelitian Kuantitatif*
- Fu Ruomei, Agustian. (2017). *Bahasa Mandarin Sistem 52M*. Jakarta: Kesaint Blanc.
- Hansen, PE, & Sutandi, S. (2022). *Analisis perbandingan pembelajaran offline dan online empat keterampilan bahasa Mandarin di salah satu perguruan tinggi swasta di Jawa Barat*. PRASI , 17 (02), 135-152.
- Hamdi, Bahrudin. (2014). *Metode Penelitian Kuantitatif Aplikasi Dalam Pendidikan*. Yogyakarta. Deepublish.
- Ina, Septy, Siti, Fika. (2021). *Analisis Validitas, Reliabilitas, Tingkat Kesulitan Dan Daya Pembeda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan*.
- Magdalena, I., Mahromiyati, M., & Nurkamilah, S. (2021). *Analisis instrumen tes Sebagai Alat Evaluasi Pada Mata Pelajaran Sbdp Siswa Kelas II SDN Duri Kosambi 06 Pagi Nusantara*.
- Mustarih, N., & Achmad, A. K. (2023, July). *Analisis Indikator Tes Kemahiran Bahasa Mandarin Tingkat Menengah: Hanyu Shuiping Kasoshi 汉语水平考试 (HSK 4)*. In Seminar Nasional Dies Natalis 62 (Vol. 1, pp. 45-58).
- Moh Soheh, Ahmad. (2020). *Evaluasi Pembelajaran Dalam Konteks Fungsi, Tujuan, dan Manfaat Yang Dilakukan Oleh Guru*.
- Prasetyaningtyas, H. (2023, January). *Analisis Buku Teks Pelajaran Bahasa Mandarin Jenjang SMA*. In *Journal International Seminar on Languages, Literature, Arts, and Education (ISLLAE)* (Vol. 5, No. 1, pp. 19-29).
- Prabowo, J., & Sutandi, S. (2023). *Analisis Latihan Keterampilan Menyimak Dan Berbicara Dalam Buku “Chinese Made Easy I Edisi Kedua”*. PRASI, 18(02), 205-222.
- Rukajat, Ajat. (2018). *Teknik Evaluasi Pembelajaran*. Yogyakarta. Deepublish.
- Setiyo, S. A., Anggraeni, A., Supriatnaningsih, R., & Marsuki, R. R. (2021). *Pengembangan Media Pembelajaran Kuis Interaktif Pilihan Ganda Bahasa Mandarin Sekolah Menengah Pertama (SMP)*. Longda Xiaokan: *Journal of Mandarin Learning and Teaching*, 4(1).
- Sugiyono. (2013). *Metode Penelitian Pendidikan (Pendekatan Kuantitatif, Kualitatif, dan R&D)*. Bandung: Alfabeta.
- Sherly, S., Dharma, E., & Purba, S. (2020). *Pemanfaatan Media Pembelajaran Untuk Meningkatkan Kemampuan Siswa Dalam Belajar Bahasa Mandarin Di Smp Swasta Sultan Agung Pematang Siantar*. *Proceeding Umsurabaya*.
- Sudjana, Nana. (2011). *Penilaian Hasil Belajar Mengajar*. Bandung: PT Raja Grafindo Persada.
- Tim Direktorat Pembinaan SMP. (2017). *Panduan Penilaian Oleh Pendidik & Satuan Pendidikan SMP*.
- Unisba. (2015). *Buku Pedoman Penilaian & Pembuatan Soal*.
- Widiyanto, Joko. (2018). *Evaluasi Pembelajaran: Sesuai dengan Kurikulum 2013*. Madiun. UNIPMA PRESS.