

## **BAB II**

### **LANDASAN TEORI**

#### **A. Deskripsi Teori**

##### **1. Evaluasi soal sumatif Bahasa Indonesia**

Menurut Arikunto (2019) penilaian adalah rangkaian kegiatan untuk memperoleh, menganalisis, dan menafsirkan data tentang proses dan hasil belajar peserta didik. Soal sumatif digunakan untuk mengukur pencapaian belajar setelah satu periode tertentu. Permendikbud No.23 Tahun 2016 menjelaskan bahwa penilaian sumatif dilakukan pada akhir pembelajaran untuk menentukan capaian kompetensi peserta didik. Arikunto (2019:72) menyatakan bahwa suatu instrumen dikatakan valid apabila mampu mengukur apa yang hendak diukur. Tetapi tiga jenis validitas yang digunakan dalam penelitian ini:

- a. Validitas isi : Validitas isi memastikan bahwa isi soal sesuai dengan cakupan materi. Menurut Nitko & Brookhart (2014:45) validasi isi mengacu pada seberapa baik isi tes mewakili domain hasil pembelajaran yang di harapkan.
- b. Validitas konstruksi : Validitas konstruksi menilai apakah sebuah soal menggambarkan kemampuan yang diukur. Menurut Crocker & Algina (2008:231) validitas konstruk adalah tingkat dimana sebuah tes mengukur konstruk teoritis yang diklamnya untuk diukur.

- c. Validitas bahasa : Validitas bahasa menekankan kejelasan dan ketepatan penggunaan bahasa dalam soal. Menurut Suharsimi Arikunto (2019:95) menyatakan bahwa bahasa soal harus komunikatif, tidak menimbulkan tafsiran ganda dan menggunakan bahasa baku.

Menurut Haladyna (2004:56) tingkat kesulitan item menunjukkan proporsi siswa yang menjawab item dengan benar. Daya pembeda menunjukkan kemampuan soal membedakan siswa pandai dan kurang pandai. Menurut Ebel & Frisbie (1991:79) indeks diskriminasi mencerminkan seberapa baik suatu item membedakan antara kinerja tinggi dan rendah. Haladyna (2004:101) menjelaskan bahwa pengalih perhatian yang baik dan menarik beberapa peserta ujian dan berfungsi untuk meningkatkan kualitas item.

## **2. Perbandingan kinerja AI ChatGPT dan DeepSeek**

Bhandari (2024) dalam studi IRT menyatakan baik parameter kesulitan maupun diskriminasi dari item yang dihasilkan ChatGPT berbeda secara signifikan dan secara statistik dari item yang ditulis manusia. Artificial Intelligence dalam pendidikan (ChatGPT dan DeepSeek). Menurut Rustamaji dan Hidayat (2023) model bahasa besar (LLM) seperti ChatGPT efektif digunakan sebagai alat pembuat dan analisis soal dalam pendidikan karena kemampuannya dalam memahami dan mensintesis teks secara otomatis. DeepSeek (diadaptasi dari dokumentasi resminya, 2024) DeepSeek berbasis retrieval-augmented generation yang dapat

digunakan untuk pemrosesan dokumen dan pembuatan soal dalam konteks pendidikan dengan hasil yang kompetitif.

### 3. Capaian Pembelajaran dan Tujuan Pembelajaran di kelas X-8 SMAN 7 Kediri

Dalam konteks Bahasa Indonesia di SMAN 7 Kediri, soal sumatif guna untuk mengukur kemampuan peserta didik menggunakan beberapa elemen yaitu membaca, memahami, menulis, memirsa, dan berbicara, peserta didik juga harus mahir dalam menganalisis teks dan menggunakan bahasa dengan tepat. Adapun capaian pembelajaran dan tujuan pembelajaran di Sekolah Menengah Atas Negeri 7 Kediri ialah:

#### a. Capaian Pembelajaran

| Elemen                         | Capaian Pembelajaran                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|--------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Menyimak                       | Peserta didik mampu mengevaluasi dan mengkreasi informasi berupa gagasan, pikiran, perasaan, pandangan, arahan atau pesan yang akurat dari menyimak berbagai tipe teks dalam bentuk monolog, dialog, dan gelar wicara.                                                                                                                                                                                                                                                                                            |
| Membaca dan Memirsa            | Peserta didik mampu mengevaluasi informasi berupa gagasan, pikiran, pandangan, arahan atau pesan dari berbagai tipe teks visual dan audiovisual untuk menemukan makna yang tersurat dan tersirat. Peserta didik mampu menginterpretasi informasi untuk mengungkapkan gagasan dan perasaan simpati, peduli, empati dan/atau pendapat pro/kontra dari teks visual dan audiovisual secara kreatif. Peserta didik mampu menggunakan sumber lain untuk menilai akurasi dan kualitas data serta membandingkan isi teks. |
| Berbicara dan Mempresentasikan | Peserta didik mampu mengolah dan menyajikan gagasan, pikiran, pandangan, arahan atau pesan untuk tujuan pengajuan usul, perumusan masalah, dan solusi dalam bentuk monolog, dialog, dan gelar wicara secara logis, runtut, kritis, dan kreatif. Peserta didik mampu mengkreasi ungkapan sesuai dengan norma kesopanan dalam berkomunikasi. Peserta didik mampu berkontribusi lebih aktif dalam diskusi dengan mempersiapkan materi diskusi, melaksanakan tugas dan fungsi dalam diskusi.                          |

|         |                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|---------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|         | Peserta didik mampu mengungkapkan kepedulian secara kreatif dalam bentuk teks fiksi dan nonfiksi multimodal.                                                                                                                                                                                                                                                                                                                                       |
| Menulis | Peserta didik mampu menulis gagasan, pikiran, pandangan, arahan atau pesan tertulis untuk berbagai tujuan secara logis, kritis, dan kreatif dalam bentuk teks informasional dan/atau fiksi. Peserta didik mampu menulis teks eksposisi hasil penelitian dan teks fungsional dunia kerja. Peserta didik mampu mengalihwahkan satu teks ke teks lainnya. Peserta didik mampu menerbitkan hasil tulisan di media cetak, elektronik, dan/atau digital. |

Tabel 2.1 Capaian Pembelajaran 1

## b. Tujuan Pembelajaran

| No. | Tujuan Pembelajaran                                                                                                                                                                                                                            |
|-----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1.  | Pelajar menilai ide pokok dan ide perinci serta merekonstruksi menjadi informasi dengan kata-kata sendiri setelah menyimak teks nonfiksi dalam gelar wicara, yaitu ceramah/pidato.                                                             |
| 2.  | Pelajar menganalisis kata-kata jarang muncul dan kata-kata baru dalam teks eksplanasi dengan strategi kosakata (menggunakan kamus, tesaurus, atau strategi prediksi), dan menggunakan kata-kata tersebut dalam konteks spesifik yang dikenali. |
| 3.  | Pelajar menganalisis kata-kata jarang muncul dan kata-kata baru dalam teks eksplanasi dengan strategi kosakata (menggunakan kamus, tesaurus, atau strategi prediksi), dan menggunakan kata-kata tersebut dalam konteks spesifik yang dikenali. |
| 4.  | Pelajar menilai pesan setelah menyimak teks sastra lisan populer (prosa, puisi, atau drama) yang berbentuk monolog atau dialog, dengan kata-kata sendiri secara kritis dan reflektif.                                                          |
| 5.  | Pelajar menilai dan memprediksi tindakan tokoh dalam teks rekon dari teks visual dan audiovisual.                                                                                                                                              |
| 6.  | Pelajar mampu menulis gagasan dalam bentuk teks naratif melalui pengembangan kerangka alur cerita.                                                                                                                                             |
| 7.  | Pelajar menyimpulkan gagasan penulis dari beberapa teks artikel ilmiah remaja.                                                                                                                                                                 |
| 8.  | Pelajar mampu menulis teks fungsional dunia kerja, seperti surat niaga atau surat lamaran pekerjaan berdasarkan iklan ataupun tanpa iklan.                                                                                                     |
| 9.  | Pelajar menafsirkan alasan atau makna karakterisasi dan alur cerita pada teks rekon dengan kata-kata sendiri.                                                                                                                                  |
| 10. | Pelajar menulis teks eksposisi dengan informasi yang akurat dan merujuk pada sumber-sumber informasi yang valid terkait topik yang dikenali.                                                                                                   |
| 11. | Pelajar memvalidasi gagasan dan langkah-langkah dari kegiatan menyimak teks prosedur dalam bentuk gelar wicara.                                                                                                                                |
| 12. | Pelajar menulis gagasan dalam bentuk teks argumentatif terkait sebuah fenomena yang telah dikenali dengan dalil dan bukti yang berasal dari pengamatan, pengalaman, dan rujukan yang diketahui.                                                |
| 13. | Pelajar membuktikan akurasi penggambaran keragaman masyarakat dalam teks deskriptif dengan tema tertentu.                                                                                                                                      |
| 14. | Pelajar menyajikan gagasan dalam bentuk usul/proposal kegiatan secara lisan dengan menggunakan kata-kata sendiri dalam bentuk                                                                                                                  |

|     |                                                                                                                                                                                                                                                                         |
|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|     | monolog, dialog, dan gelar wicara.                                                                                                                                                                                                                                      |
| 15. | Pelajar menulis esai tentang fenomena sains, sosial, humaniora yang dikenali secara padu dan koheren.                                                                                                                                                                   |
| 16. | Pelajar memberi argumentasi dan saran dari kegiatan menyimak teks laporan hasil pengamatan atau percobaan dalam bentuk monolog, dialog, dan gelar wicara.                                                                                                               |
| 17. | Pelajar memproduksi gagasan dalam kegiatan debat.                                                                                                                                                                                                                       |
| 18. | Pelajar menilai akurasi dan kualitas penyajian data terhadap beragam teks laporan penelitian dengan kata-kata sendiri secara kritis dan reflektif.                                                                                                                      |
| 19. | Pelajar mengidentifikasi penyebab masalah atau kejadian, hubungan sebab akibat yang melatari, merumuskan masalah utama, dan mengusulkan solusi terkait dengan teks eksplanasi dalam sebuah diskusi.                                                                     |
| 20. | Pelajar menuliskan hasil riset dengan metodologi sederhana terkait topik yang dikenali dengan mengutip sumber rujukan.                                                                                                                                                  |
| 21. | Pelajar mempresentasikan ide atau pendapat tentang topik yang dikenali, secara elaboratif dan rinci, yang dilengkapi argumen pendukung yang dapat dipertanggungjawabkan, dan menggunakan intonasi dan metode presentasi yang mengundang perhatian atau minat pendengar. |
| 22. | Pelajar terampil dan terbiasa menulis indah (puisi atau prosa) berdasarkan pengetahuan, pengalaman, pengamatan, atau gagasan sendiri, terkait dengan topik yang dikenali yang bisa diterbitkan di media cetak maupun digital.                                           |

Tabel 2.2 Tujuan Pembelajaran 1

## B. Kerangka Berpikir

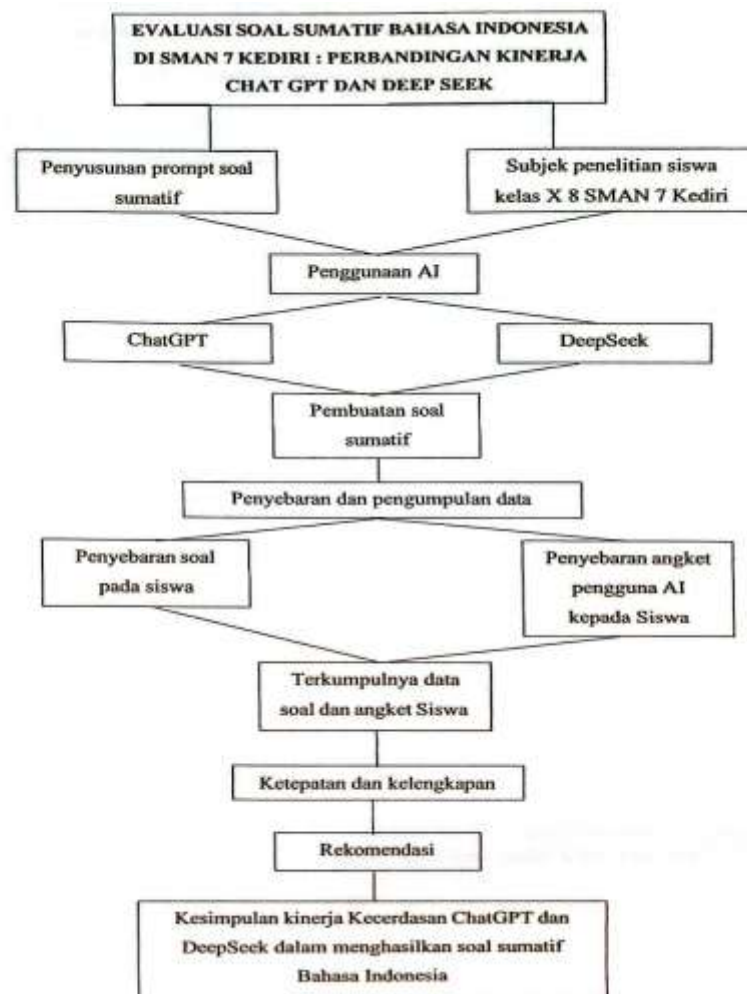
### 1. Tabel kerangka berpikir

| Tahapan Penelitian         | Uraian                                                                                                                                                                                                                                                          |
|----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Permasalahan               | Kualitas soal sumatif Bahasa Indonesia perlu dievaluasi untuk mengetahui kesesuaiannya dengan kaidah penyusunan instrumen penilaian. Perkembangan kecerdasan buatan menawarkan alternatif evaluasi soal yang cepat dan sistematis melalui ChatGPT dan DeepSeek. |
| Objek Penelitian           | Soal sumatif Bahasa Indonesia yang digunakan di SMAN 7 Kediri                                                                                                                                                                                                   |
| Instrumen Evaluasi         | Kriteria evaluasi soal meliputi kesesuaian dengan Capaian Pembelajaran, Tujuan Pembelajaran, tingkat kognitif, konstruksi soal, dan penggunaan bahasa                                                                                                           |
| Proses Analisis            | Prompt yang saya di upload pada kecerdasan buatan yaitu ChatGPT dan DeepSeek berdasarkan kriteria prompt yang telah di tetapkan.                                                                                                                                |
| Variabel yang Dibandingkan | Ketetapan hasil soal, kelengkapan hasil soal yang sesuai dengan prompt yang di unggah, kemampuan kecerdasan buatan                                                                                                                                              |

|                    |                                                                                                                                                                                  |
|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                    | dalam menghasilkan soal yang sesuai dengan prompt yang di unggah, dan kualitas soal yang di hasilkan oleh ChatGPT dan DeepSeek.                                                  |
| Hasil Perbandingan | Diperoleh informasi mengenai persamaan dan perbedaan kinerja kecerdasan buatan ChatGPT dan DeepSeek dalam menghasilkan soal sumatif kelas X 8 Bahasa Indonesia di SMAN 7 Kediri. |
| Kesimpulan         | Menentukan platform yang menghasilkan soal lebih akurat, komprehensif, dan sesuai dengan instrumen prompt yang di unggah juga kesesuaian bahasa yang efektif.                    |

Tabel 2.3 Kerangka Berpikir 1

## 2. Bagan kerangka berpikir



Gambar 2.1 Kerangka Berpikir 1

### 3. Hipotesis Penelitian

Hipotesis penelitian merupakan jawaban sementara terhadap rumusan masalah yang diajukan, yang kebenarannya akan dibuktikan secara empiris melalui analisis statistik. Berdasarkan kerangka berpikir yang telah dipaparkan, hipotesis dalam penelitian ini dibagi menjadi hipotesis utama dan hipotesis spesifik sebagai berikut:

#### a. Hipotesis Utama

Hipotesis utama ini berkaitan dengan perbandingan kinerja secara menyeluruh antara dua model kecerdasan buatan (Large Language Model) yang diuji.

- 1) H0 (Hipotesis Nol): Tidak terdapat perbedaan yang signifikan antara ChatGPT dan DeepSeek dalam hal validitas (isi dan empiris) serta kualitas (tingkat kesukaran, daya pembeda, dan efektivitas pengecoh) soal sumatif Bahasa Indonesia di SMAN 7 Kediri.
- 2) H1 (Hipotesis Alternatif): Terdapat perbedaan yang signifikan antara ChatGPT dan DeepSeek dalam hal validitas (isi dan empiris) serta kualitas (tingkat kesukaran, daya pembeda, dan efektivitas pengecoh) soal sumatif Bahasa Indonesia di SMAN 7 Kediri.

#### b. Hipotesis Spesifik

Hipotesis spesifik merinci perbedaan pada setiap parameter kualitas butir soal yang diukur dalam penelitian:

1) Validitas Isi (Penilaian Ahli)

H01: Tidak terdapat perbedaan signifikan pada nilai rata-rata validitas isi oleh ahli antara soal yang dianalisis/dihasilkan oleh ChatGPT dan DeepSeek.

H11: Terdapat perbedaan signifikan pada nilai rata-rata validitas isi oleh ahli antara soal yang dianalisis/dihasilkan oleh ChatGPT dan DeepSeek.

2) Validitas Empiris (Korelasi Item-Total)

H02: Tidak terdapat perbedaan signifikan pada nilai korelasi item-total (validitas empiris) antara hasil analisis ChatGPT dan DeepSeek.

H12: Terdapat perbedaan signifikan pada nilai korelasi item-total (validitas empiris) antara hasil analisis ChatGPT dan DeepSeek.

3) Tingkat Kesukaran ( $p$ )

H03: Tidak terdapat perbedaan signifikan pada rata-rata indeks tingkat kesukaran antara soal yang diproses oleh ChatGPT dan DeepSeek.

H13: Terdapat perbedaan signifikan pada rata-rata indeks tingkat kesukaran antara soal yang diproses oleh ChatGPT dan DeepSeek.

4) Daya Pembeda ( $D$ )

H04: Tidak terdapat perbedaan signifikan pada nilai rata-rata indeks daya pembeda antara hasil analisis ChatGPT



dan DeepSeek.

H14: Terdapat perbedaan signifikan pada nilai rata-rata indeks daya pembeda antara hasil analisis ChatGPT dan DeepSeek.

#### 5) Efektivitas Pengecoh (Distraktor)

H05: Tidak terdapat perbedaan proporsi pengecoh efektif per butir soal antara hasil analisis ChatGPT dan DeepSeek.

H15: Terdapat perbedaan proporsi pengecoh efektif per butir soal antara hasil analisis ChatGPT dan DeepSeek.

#### 6) Keterpaduan dengan Kompetensi Dasar (KD)

H06: Tidak terdapat perbedaan skor keterpaduan butir soal dengan Kompetensi Dasar (berdasarkan penilaian ahli) antara model ChatGPT dan DeepSeek.

H16: Terdapat perbedaan skor keterpaduan butir soal dengan Kompetensi Dasar (berdasarkan penilaian ahli) antara model ChatGPT dan DeepSeek.

Pengujian Statistik: Hipotesis spesifik H01 hingga H04 serta H06 diuji menggunakan uji beda rata-rata (seperti Independent Sample t-Test jika data berdistribusi normal, atau Mann-Whitney U Test jika tidak normal). Uji Proporsi: Khusus untuk hipotesis H05 (efektivitas pengecoh), karena berkaitan dengan proporsi/kategori, pengujian yang lebih tepat sering kali menggunakan uji Chi-Square. Taraf Signifikansi: Saya pastikan

dalam Bab III (Metode Penelitian) menyebutkan bahwa seluruh hipotesis ini akan diuji pada taraf signifikansi  $\alpha = 0,05$ .