



UNIVERSITAS INDONESIA

**PENYUSUNAN SKEMA PENILAIAN
UNTUK MENILAI TULISAN SISWA DI KELAS MENULIS
BAHASA INDONESIA UNTUK PENUTUR ASING
TINGKAT MADYA**

TESIS

**HARRY PURNAMA
NPM 0906587092**

**FAKULTAS ILMU PENGETAHUAN BUDAYA
PROGRAM STUDI LINGUISTIK
DEPOK
JULI 2012**



UNIVERSITAS INDONESIA

**PENYUSUNAN SKEMA PENILAIAN
UNTUK MENILAI TULISAN SISWA DI KELAS MENULIS
BAHASA INDONESIA UNTUK PENUTUR ASING
TINGKAT MADYA**

TESIS

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Magister
Humaniora**

**HARRY PURNAMA
NPM 0906587092**

**FAKULTAS ILMU PENGETAHUAN BUDAYA
PROGRAM STUDI ILMU LINGUISTIK
DEPOK
JULI 2012**

SURAT PERNYATAAN BEBAS PLAGIARISME

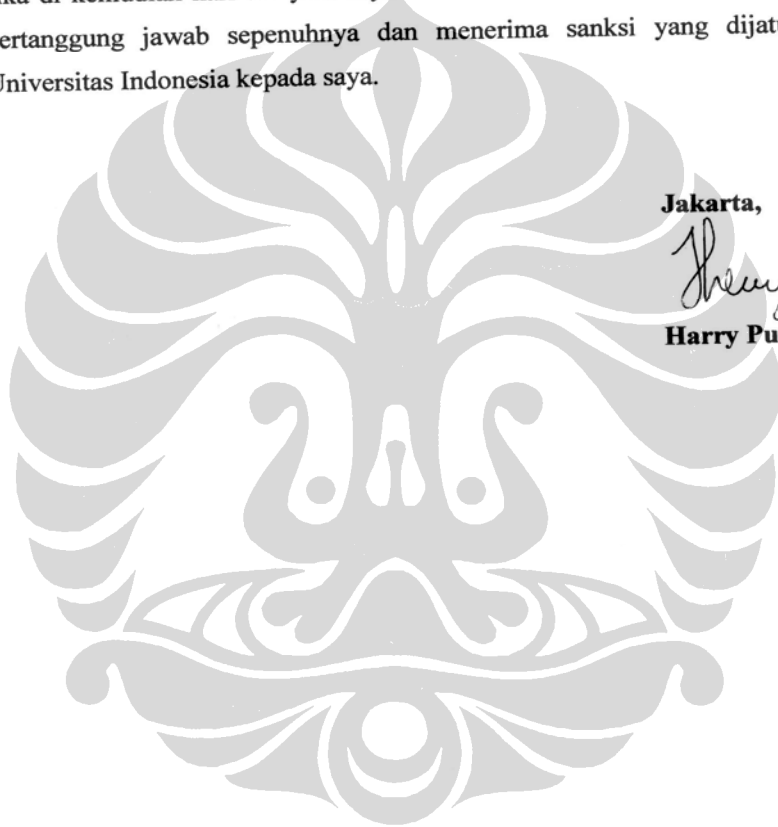
Saya yang bertanda tangan di bawah ini dengan sebenarnya menyatakan bahwa tesis ini saya susun tanpa tindakan plagiarisme sesuai dengan peraturan yang berlaku di Universitas Indonesia.

Jika di kemudian hari ternyata saya melakukan tindakan Plagiarisme, saya akan bertanggung jawab sepenuhnya dan menerima sanksi yang dijatuhkan oleh Universitas Indonesia kepada saya.

Jakarta,



Harry Purnama



HALAMAN PERNYATAAN ORISINALITAS

**Tesis ini adalah hasil karya saya sendiri,
dan semua sumber baik yang dikutip maupun dirujuk
telah saya nyatakan dengan benar.**

Nama : Harry Purnama

NPM : 0906587092

Tanda Tangan : 

Tanggal : 16 Juli 2012

HALAMAN PENGESAHAN

Tesis yang diajukan oleh :
nama : Harry Purnama
NPM : 0906587092
Program Studi : Ilmu Linguistik
judul : Penyusunan Skema Penilaian Untuk Menilai
Tulisan Siswa Di Kelas Menulis Bahasa Indonesia
Untuk Penutur Asing Tingkat Madya

ini telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Magister Humaniora pada Program Studi Ilmu Linguistik Fakultas Ilmu Pengetahuan Budaya, Universitas Indonesia.

DEWAN PENGUJI

Pembimbing : Sisilia Setiawati Halimi, Ph.D

(.....)

Penguji : Dr. F.X. Rahyono

(.....)

Penguji : Dr. Grace Wiradisastira

(.....)

Ditetapkan di : Depok
tanggal : 16 Juli 2012

oleh
Dekan
Fakultas Ilmu Pengetahuan Budaya
Universitas Indonesia

Dr. Bambang Wibawarta
NIP 19651023 199003 1 002

UCAPAN TERIMA KASIH

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa karena atas berkat dan rahmat-Nya penulis dapat menyelesaikan tesis ini. Penulis menyadari bahwa tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan tesis ini, sangatlah sulit bagi penulis untuk menyelesaikan tesis ini dengan baik

Oleh karena itu, dengan segala kerendahan hati, penulis ingin mengucapkan terima kasih kepada Ibu Sisilia Setiawati Halimi, Ph.D selaku dosen pembimbing yang telah membimbing dan mengarahkan penulis dengan penuh kesabaran dan ketelitian selama tiga semester ini. Tanpa bantuan beliau, mungkin tesis ini tidak akan memiliki isi yang berbobot. Semoga Tuhan Yang Maha Esa melimpahkan kesehatan dan keberkahan kepada beliau.

Terima kasih juga penulis ucapkan kepada Bapak Dr. F.X. Rahyono dan Ibu Dr. Grace Wiradisastra selaku penguji atas dukungan dan kesediaannya memberikan kritik dan saran untuk memperbaiki tesis ini agar mencapai hasil yang maksimal. Penulis juga mengucapkan terima kasih kepada para dosen program studi Linguistik di Fakultas Ilmu Pengetahuan Budaya Universitas Indonesia, khususnya kelas Pengajaran Bahasa Asing, karena telah mengajarkan ilmu yang bermanfaat kepada penulis selama masa perkuliahan. Semoga Tuhan YME melimpahkan pahala yang senantiasa mengalir atas apa yang telah mereka berikan itu.

Penulis juga ingin mengucapkan terima kasih kepada Ibu Sri Munawarah sebagai pembimbing akademik. Terima kasih atas pinjaman buku dan dukungan yang telah diberikan selama penulis mengerjakan tesis ini.

Penulis juga ingin mengucapkan terima kasih kepada Ibu Nunung, Nindira, Yasmin, Desril, dan Ika yang di tengah-tengah kesibukannya mau meluangkan waktu untuk membantu penulis dalam menyelesaikan pembuatan tesis ini. Tak lupa penulis juga mengucapkan terima kasih kepada pengajar kelas menulis tingkat madya yang lain karena telah direpotkan oleh penulis untuk mengumpulkan tulisan dari setiap kelas, yaitu Ibu Niken, Ibu Pris, dan Ibu Erni. Kepada Lembaga Bahasa Internasional, terutama kepada Ibu Irzanti dan Ibu Dwi

selaku Ketua BIPA, penulis juga mengucapkan terima kasih karena telah mengizinkan penulis untuk melakukan penelitian di lembaga bahasa tersebut. Juga terima kasih kepada Mbak Nana untuk tukar ilmunya, semoga cepat lulus, mbak.

Ucapan terima kasih juga penulis haturkan kepada Ayahanda Sodikin dan Ibunda Sri Maimun Farid karena telah memberikan dukungan yang luar biasa bagi penulis untuk segera menyelesaikan tesis ini. Penulis juga ingin meminta maaf karena tidak dapat menyelesaikan pendidikan S2 ini tepat waktu.

Terima kasih juga penulis ucapkan kepada teman-teman seperjuangan di Program Studi Linguistik angkatan 2009 yang telah memberikan persahabatan dan kebersamaan selama masa perkuliahan. Kepada teman seperjuangan Haje, Rita, Stella, Pak Donny, Inonk, Indah, Neneng, dan Dhuha, sampai bertemu lagi di Balairung pada perayaan wisuda. Untuk teman-teman S2 lain yang tidak dapat disebutkan satu per satu namanya, penulis juga mengucapkan terima kasih.

Yang terakhir penulis ingin mengucapkan terima kasih kepada teman-teman di Program Studi Sastra Indonesia, terutama angkatan 2003, yang tidak dapat disebutkan satu per satu. Terima kasih karena tidak bosan-bosannya menanyakan kepada penulis kapan lulus dengan maksud untuk “merayakannya”. Juga kepada teman-teman di Majalah *VENUE*, terima kasih karena memberikan dukungan kepada penulis, terutama kelonggaran waktu untuk bekerja. Penulis juga mengucapkan terima kasih kepada semua pihak yang telah membantu yang tidak dapat disebutkan namanya di lembar Ucapan Terima Kasih ini. Akhir kata, penulis berharap Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu.

Jakarta, 15 Juli 2012



Harry Purnama

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Indonesia, saya yang bertanda tangan di bawah ini:

Nama : Harry Purnama
NPM : 0906587092
Program Studi : Ilmu Linguistik
Departemen : Linguistik
Fakultas : Ilmu Pengetahuan Budaya
Jenis karya : Tesis

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Indonesia **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty Free Right*)** atas karya ilmiah saya yang berjudul:

Penyusunan Skema Penilaian Untuk Menilai Tulisan Siswa Di Kelas Menulis Bahasa Indonesia Untuk Penutur Asing Tingkat Madya

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Indonesia berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Depok
Pada tanggal : 16 Juli 2012
Yang menyatakan



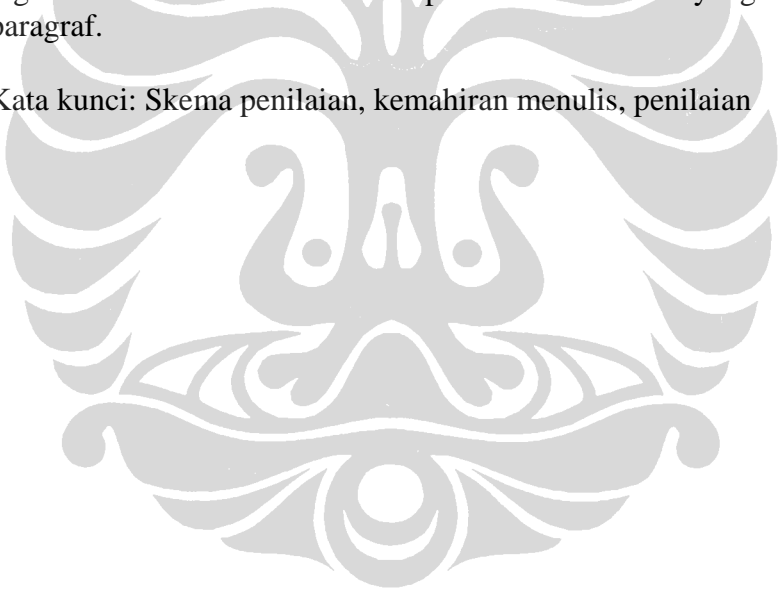
(Harry Purnama)

ABSTRAK

Nama : Harry Purnama
Program studi : Ilmu Linguistik
Judul : Penyusunan Skema Penilaian untuk Menilai Tulisan Siswa di
Kelas Menulis Bahasa Indonesia untuk Penutur Asing Tingkat
Madya

Penelitian ini bertujuan membuat skema penilaian untuk kelas menulis tingkat madya di lembaga pengajaran bahasa Indonesia untuk penutur asing. Skema penilaian yang coba dibuat adalah yang praktis digunakan dan menghasilkan penilaian yang andal. Skema dirancang dengan menggunakan langkah-langkah penyusunan skema penilaian yang dikemukakan oleh Mertler (2001). Pengujian skema penilaian dilakukan dengan bantuan tiga partisipan yang menilai tulisan yang sama. Sampel untuk pengujian skema penilaian diambil dari hasil ujian tengah semester dan ujian akhir semester dari enam kelas menulis bahasa Indonesia untuk orang asing tingkat madya periode Januari-Mei 2011. Data dari hasil penilaian itu kemudian dianalisis menggunakan ANOVA. Hasil dari pengujian skema penilaian menunjukkan bahwa skema penilaian ini lebih cocok digunakan untuk menilai esai daripada menilai tulisan yang hanya terdiri dari satu paragraf.

Kata kunci: Skema penilaian, kemahiran menulis, penilaian

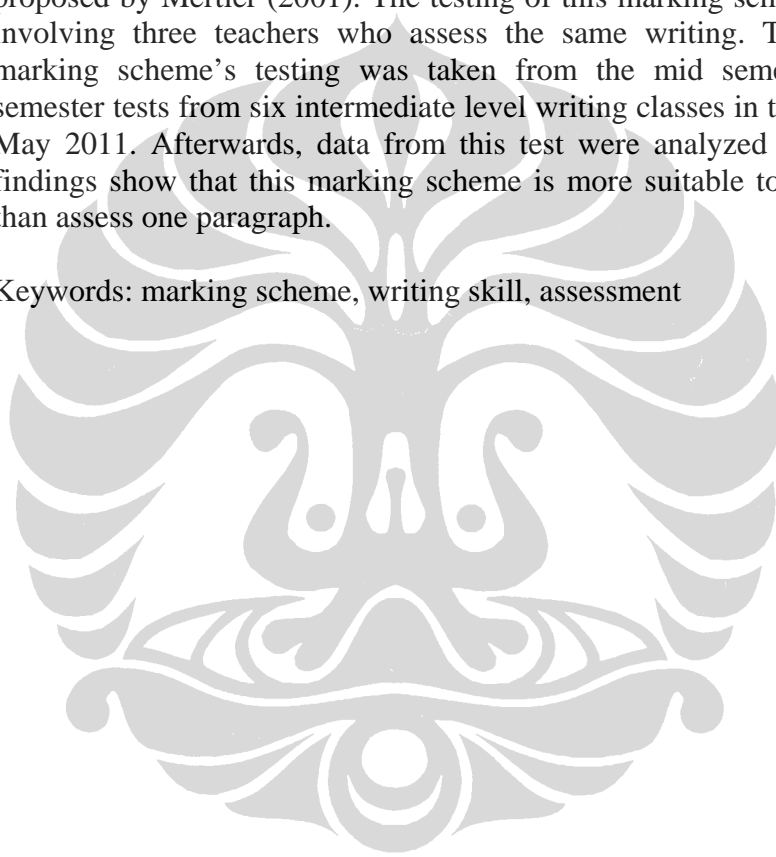


ABSTRACT

Name : Harry Purnama
Study program: Linguistics
Judul : Developing Marking Scheme for Assessing Writing of the
Intermediate Class in Indonesian Languages Learning Institution
for Foreign Speakers

The purpose of this study is to create a marking scheme for an intermediate writing class in Indonesian languages learning institution for foreign speakers. The marking scheme that I try to create is the one that is easy to use and yield reliable scores. This marking scheme is created by following the procedures proposed by Mertler (2001). The testing of this marking scheme is carried out by involving three teachers who assess the same writing. The samples for this marking scheme's testing was taken from the mid semester tests and final semester tests from six intermediate level writing classes in the period of January-May 2011. Afterwards, data from this test were analyzed using ANOVA. The findings show that this marking scheme is more suitable to assess essays rather than assess one paragraph.

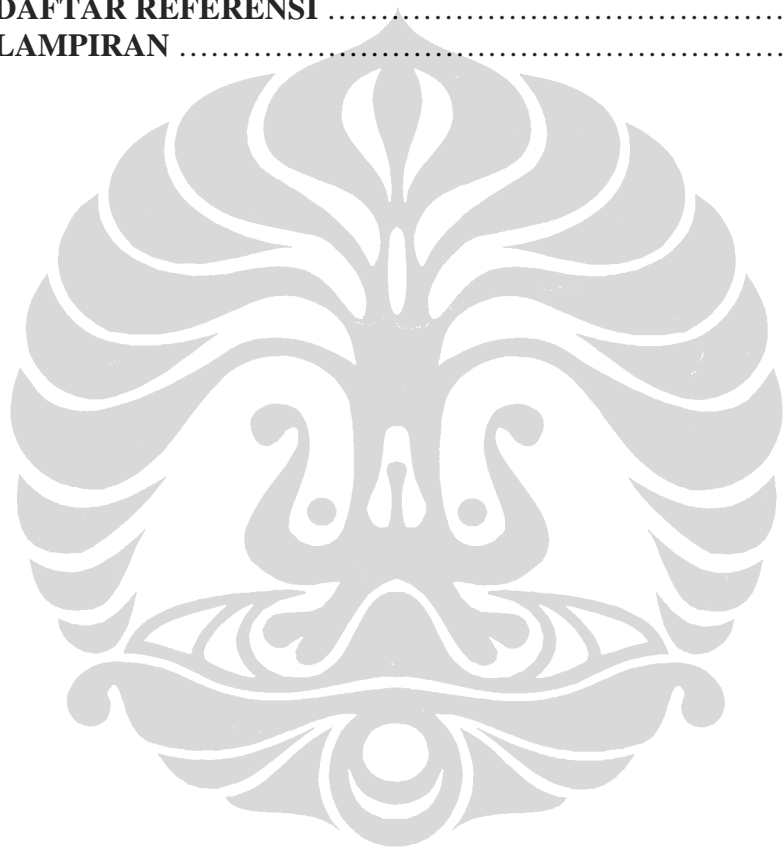
Keywords: marking scheme, writing skill, assessment



DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERNYATAAN BEBAS PLAGIARISME	ii
HALAMAN PERNYATAAN ORISINALITAS	iii
LEMBAR PENGESAHAN	iv
UCAPAN TERIMA KASIH	v
LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH	vii
ABSTRAK	viii
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR LAMPIRAN	xiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Pertanyaan Penelitian	9
1.3 Tujuan Penelitian	9
1.4 Ruang Lingkup Penelitian	9
1.5 Manfaat Penelitian	10
BAB 2 TINJAUAN PUSTAKA	11
2.1 Penelitian Terdahulu	11
2.2 Kemahiran Menulis	17
2.3 Tes dan Penilaian	20
2.4 Jenis Tugas Menulis	23
2.4.1 Jenis Tes Tulisan <i>Imitative</i>	24
2.4.2 Jenis Tes Tulisan <i>Intensive</i>	25
2.4.3 Jenis Tes Tulisan <i>Responsive</i> dan <i>Extensive</i>	27
2.5 Prosedur Penilaian	29
2.5.1 Jenis Skala Penilaian	30
2.5.1.1 Skala <i>Primary Trait</i>	30
2.5.1.2 Skala Holistik	31
2.5.1.3 Skala Analitik	33
2.6 Prinsip Praktis, Andal, dan Sahih	37
2.7 Landasan Teori	40
BAB 3 METODE PENELITIAN	42
3.1 Populasi dan Sampel	42
3.2 Partisipan Penelitian	43
3.3 Teknik Pengumpulan Data	44
3.4 Data	45
3.5 Penyusunan Skema Penilaian	45
3.5.1 Penentuan Tujuan Pembuatan Skema Penilaian	47
3.5.2 Penentuan Skema Penilaian dan Kriteria Penilaian	48
3.5.3 Penentuan Skala Penilaian dan Deskripsi Naratif	50
3.5.4 Percobaan Skema Penilaian	52
3.5.5 Perbaikan Skema Penilaian	53
3.5.6 Proses Pengujian Skema Penilaian	55

BAB 4 ANALISIS DAN PEMBAHASAN	57
4.1 Mean	59
4.2 Uji Normalitas	60
4.3 Uji Homogenitas Varian	64
4.4 Uji Perbandingan	65
4.5 Pembahasan	69
4.6 Tindak Lanjut	73
BAB 5 PENUTUP	75
5.1 Simpulan	75
5.2 Saran	76
DAFTAR REFERENSI	78
LAMPIRAN	82

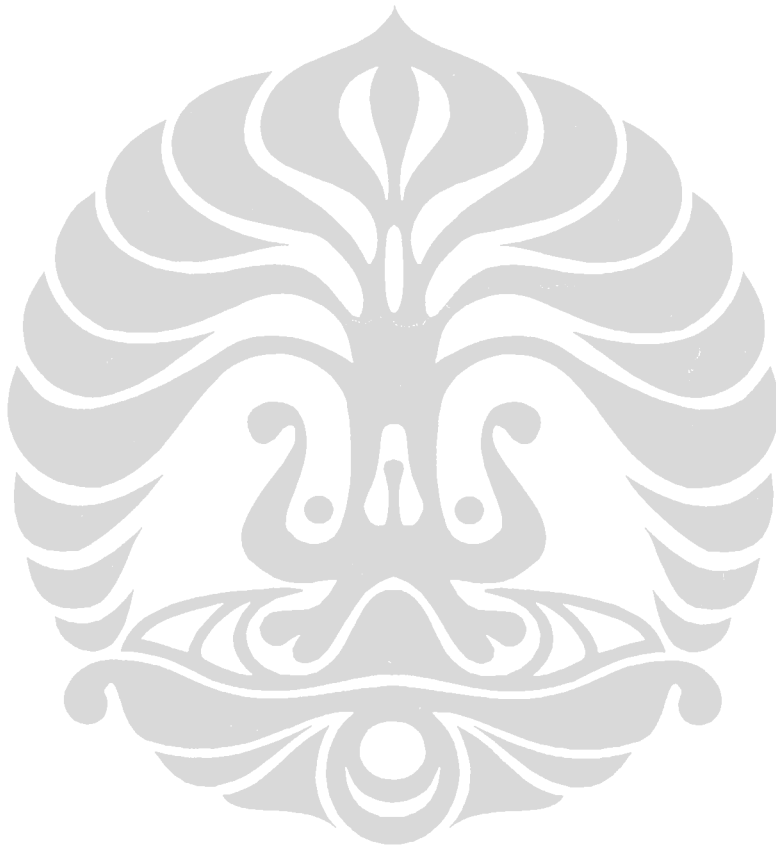


DAFTAR TABEL

Tabel 2.1	Skala Holistik	33
Tabel 2.2	Skala Analitik	36
Tabel 3.1	Bentuk Akhir Skema Penilaian yang Diadaptasi dari Jacobs	54
Tabel 4.1	Data Partisipan 1	58
Tabel 4.2	Data Partisipan 2	58
Tabel 4.3	Data Partisipan 3	58
Tabel 4.4	Nilai Rata-Rata Ketiga Partisipan	59
Tabel 4.5	Uji Normalitas UTS	63
Tabel 4.6	Uji Normalitas UAS	64
Tabel 4.7	Uji Homogenitas Varian UTS	64
Tabel 4.8	Uji Homogenitas Varian UAS	65
Tabel 4.9	Deskriptif UTS	66
Tabel 4.10	Uji Perbandingan UTS	66
Tabel 4.11	Uji Duncan UTS	67
Tabel 4.12	Deskriptif UAS	68
Tabel 4.13	Uji Perbandingan UAS	68
Tabel 4.14	Uji Duncan UAS	69
Tabel 4.15	Skema Penilaian Adaptasi dari Jacobs yang Telah Diperbaiki	74

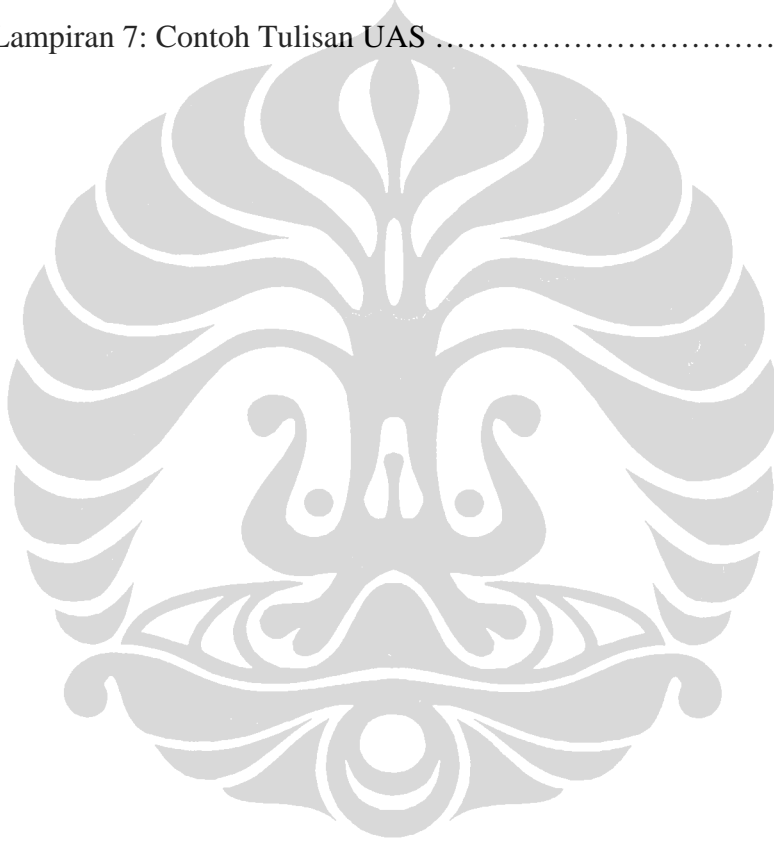
DAFTAR GAMBAR

Gambar 4.1 Uji Normalitas Skewness Kurtosis	61
---	----



DAFTAR LAMPIRAN

Lampiran 1: Borang yang Diujikan	82
Lampiran 2: Borang yang Telah Diperbaiki Hasil Wawancara	83
Lampiran 3: Transkrip Wawancara Partisipan 1	84
Lampiran 4: Transkrip Wawancara Partisipan 2	87
Lampiran 5: Contoh Borang yang telah Digunakan	88
Lampiran 6: Contoh Tulisan UTS	89
Lampiran 7: Contoh Tulisan UAS	90



BAB 1

PENDAHULUAN

1.1 Latar Belakang

Dengan terbukanya era globalisasi dan semakin majunya dunia teknologi, dunia pekerjaan dan pendidikan menjadi terbuka luas. Batas-batas antarnegara hampir tidak terlihat lagi. Masyarakat di Indonesia dengan mudah mencari suatu informasi di negara lain. Dengan bantuan teknologi yang semakin canggih, kita bisa berkomunikasi dengan orang lain di belahan dunia yang jauh. Bahkan, kita pun bisa kuliah melalui dunia maya.

Salah satu imbas dari semakin tipisnya batas antarnegara adalah banyak masyarakat Indonesia yang bepergian ke luar negeri, baik untuk berjalan-jalan, bekerja, atau untuk belajar. Begitu pula sebaliknya, banyak orang asing yang datang ke Indonesia, baik untuk belajar maupun untuk bekerja. Orang asing ini tentu saja sedikit yang sudah mengerti bahasa Indonesia. Umumnya mereka menggunakan bahasa pengantar bahasa Inggris. Orang asing yang bekerja di Indonesia tersebut tentu juga harus berkomunikasi dengan karyawan atau anak buah yang mayoritas adalah orang Indonesia. Sebagai contoh orang asing yang bekerja sebagai manajer di sebuah perusahaan tekstil harus berkomunikasi dengan pegawainya dengan bahasa Indonesia, bukan bahasa Inggris, sebab mayoritas pegawai di perusahaan tekstil adalah pekerja yang tidak memiliki pendidikan tinggi karena itu mereka tidak mengerti bahasa Inggris. Para suami yang bekerja di Indonesia biasanya membawa serta istrinya ke Indonesia. Sebagai ibu rumah tangga, para istri itu juga sangat penting untuk menguasai bahasa Indonesia, sebab di kehidupan sehari-hari mereka setidaknya harus berkomunikasi dengan sopir, dengan pembantu, atau dengan pelayan toko.

Dari contoh-contoh di atas, terlihat bahwa orang asing yang datang ke Indonesia untuk bekerja membutuhkan sesuatu yang sangat fundamental, yaitu kemampuan untuk berkomunikasi dalam bahasa Indonesia. Pada umumnya mereka ingin belajar bahasa Indonesia dengan tujuan akhir dapat berkomunikasi menggunakan bahasa Indonesia dengan lancar, bukan untuk mencari nilai yang terbaik di kelasnya. Mereka membutuhkan sarana yang dapat mengajarkan bahasa Indonesia, baik itu sebuah lembaga bahasa atau hanya belajar secara privat.

Hal tersebutlah yang mendorong beberapa lembaga pendidikan di Indonesia untuk menawarkan pengajaran bahasa Indonesia bagi penutur asing. Lembaga bahasa tersebut menawarkan beberapa pilihan studi yang dapat dipilih oleh orang asing. Misalnya, untuk ibu rumah tangga bisa mengambil kelas khusus komunikasi sehari-hari. Selain itu, bagi murid yang tidak dapat setiap hari hadir ke kelas karena terlalu sibuk, beberapa lembaga bahasa juga menawarkan memberi pelajaran secara privat yang waktu dan tempat ditentukan oleh murid.

Di Indonesia, banyak lembaga pengajaran bahasa Indonesia untuk penutur asing yang masih berusia muda. Dengan usia yang masih muda, tentu lembaga bahasa itu membutuhkan saran-saran untuk mengembangkan lembaga itu menjadi lebih baik dan sempurna dalam memberikan pelajaran bahasa Indonesia untuk penutur asing. Pengembangan itu bisa dalam hal program, kurikulum, metode dan bahan pelajaran, pengembangan kelembagaan, serta pengembangan mutu pengajar.

Bidang program dan kurikulum mencakup berbagai tujuan pengajaran bahasa Indonesia untuk penutur asing serta keterampilan yang diajarkan. Inti dari pengembangan bidang ini adalah untuk mengarahkan kepada standarisasi program dan peningkatan mutu program. Sebagaimana diketahui, berbagai program yang ada di lembaga pengajaran bahasa Indonesia untuk penutur asing mempunyai tujuan yang berbeda-beda, antara lain ada yang bertujuan untuk dapat berkomunikasi secara lisan sehari-hari, untuk dapat membaca buku atau koran, atau untuk dapat menulis secara akademis.

Metode pengajaran bahasa Indonesia untuk penutur asing berbeda dengan metode pengajaran bahasa Indonesia untuk penutur bahasa Indonesia. Pengajaran bahasa asing sudah banyak dilakukan oleh lembaga-lembaga pengajaran bahasa asing di Indonesia, seperti pengajaran bahasa Inggris, Prancis, dan Mandarin. Lembaga-lembaga pengajaran bahasa asing itu tentu memiliki metode pengajaran yang sudah mapan untuk digunakan. Oleh karena itu, sebuah lembaga pengajaran bahasa Indonesia untuk penutur asing dapat belajar dari pengalaman lembaga-lembaga itu. Pengembangan metode akan berdampak pada pengembangan bahan pelajaran dan pengembangan mutu pengajar.

Pengembangan kelembagaan merupakan kelanjutan dari pengembangan di bidang program dan kurikulum. Pengembangan lembaga melalui pengembangan

organisasi diperlukan untuk mengefektifkan tenaga pengajar serta meningkatkan program-program ekstrakurikuler, seperti program sosial budaya yang sangat diperlukan oleh para siswa. Jika sebuah lembaga pengajaran bahasa Indonesia sudah tersertifikasi dan terakreditasi, ada kemungkinan lembaga itu menjadi pilihan teratas bagi calon siswa sebagai tempat belajar.

Pengajar merupakan unsur penting dalam pengajaran bahasa Indonesia untuk penutur asing. Pada umumnya, pengajar itu tidak pernah mengalami pendidikan khusus untuk menjadi pengajar bahasa Indonesia untuk penutur asing. Oleh karena itu, program-program penataran untuk pengajar bahasa Indonesia untuk penutur asing selalu diperlukan agar mereka menguasai metode dan teknik mengajar yang sesuai untuk pengajaran bahasa Indonesia untuk penutur asing yang berbeda dari metode dan teknik pengajaran bahasa Indonesia untuk penutur bahasa Indonesia.

Walaupun proses pembelajaran memegang peranan penting untuk dapat berkomunikasi dalam bahasa Indonesia, penilaian kemahiran tersebut juga sama pentingnya, tidak terkecuali kemahiran menulis. Semakin tingginya kebutuhan untuk dapat menulis dalam bahasa Indonesia sebagai bahasa asing atau bahasa kedua, kebutuhan untuk penilaian yang sahih dan andal juga semakin besar, baik untuk penggunaan di kelas atau untuk memprediksi kesuksesan akademis.

Perkembangan penilaian performa siswa semakin menggantikan penilaian tradisional yang mengharuskan siswa menjawab pertanyaan pilihan ganda. Hal tersebut tentu berpengaruh juga terhadap cara menilai performa tersebut. Dibutuhkan suatu cara penilaian yang benar-benar dapat menilai performa siswa secara tepat. Proses penilaian merupakan salah satu unsur penting karena nilai yang diberikan oleh penilai akan digunakan untuk menarik kesimpulan terhadap si penulis atau murid. Selain itu, bagi murid, nilai yang ia terima juga menjadi tolok ukur keberhasilannya dalam mempelajari bahasa Indonesia. Dari nilai yang diterima oleh murid itu, mereka dapat mengetahui banyak hal, seperti apakah mereka sudah mengerti pokok pelajaran yang diberikan, atau apakah mereka sudah mengerti tugas yang diberikan oleh pengajar.

Sammeng (1996) mengatakan bahwa dalam program pengajaran, alat evaluasi diperlukan untuk mengukur keberhasilan peserta dalam mencapai tujuan

belajarnya maupun dalam mengukur keberhasilan program yang diselenggarakan. Dengan demikian, setidaknya diperlukan dua jenis alat evaluasi, yaitu alat evaluasi hasil belajar dan alat evaluasi program pengajaran.

Menilai tulisan siswa bukan pekerjaan yang mudah karena berbeda dengan soal berupa pilihan ganda atau pertanyaan benar-salah (*yes-no question*). Untuk soal berupa pilihan ganda atau betul-salah, menghitung nilai dari jawaban siswa tentu sangat mudah karena kriterianya sudah jelas, yakni berapa jumlah jawaban yang benar dan berapa yang salah. Sementara itu, untuk menentukan berapa nilai yang patut diberikan atas pekerjaan menulis siswa, sebuah tulisan memiliki beberapa unsur yang perlu dinilai, baik dari segi kosakata, tata bahasa, ejaan, dan lainnya. Pada akhirnya, penilai atau pengajar akan mengira-ngira berapa nilai yang harus diberikan. Tentu saja hal itu menjadi tidak sah karena tidak didukung bukti yang jelas.

Proses menilai tulisan juga memakan waktu yang lama dan membutuhkan konsistensi penilaian dari pengajar, apalagi jika ia juga mengajar di beberapa kelas berbeda dan tiap kelas memiliki murid yang banyak. Tentu hal itu mengakibatkan terbatasnya waktu yang dimiliki si pengajar untuk menilai pekerjaan siswa. Karena itu, diperlukan suatu instrumen atau skema untuk menilai tulisan siswa secara cepat dan andal.

Penggunaan skema penilaian untuk menilai sebuah tulisan tentu sangat membantu. Namun, tidak semua skema penilaian memiliki tingkat kemudahan dan keandalan dalam menilai. Juga, skema penilaian itu belum tentu akan menghasilkan nilai yang sama apabila sebuah tulisan dinilai oleh dua atau tiga penilai yang berbeda. Selain itu, alat penilaian yang akan digunakan juga harus mudah dan praktis digunakan.

Untuk menilai sebuah tulisan, alat yang paling sering digunakan adalah dengan menggunakan skema penilaian atau biasa disebut juga rubrik. Weigle (2002) dan Brown (2004) menyebutkan ada tiga jenis skema penilaian yang umum digunakan untuk menilai sebuah tulisan, yakni skala *primary trait*, skala holistik, dan skala analitik. Di antara ketiga skema penilaian itu, yang paling sering digunakan adalah skala holistik dan skala analitik.

Skala holistik memiliki kelebihan dalam hal efektivitas waktu. Skala holistik menilai sebuah tulisan berdasarkan impresi yang didapat dari penilai ketika membaca tulisan itu. Penilai hanya memberikan nilai tunggal untuk tulisan itu. Kekurangan skala holistik adalah tidak dapat memberi balikan yang efektif kepada murid dan pengajar sebab nilai yang diberikan tidak memberikan informasi kekurangan dan kelebihan tulisan yang dinilai. Sementara itu, kelebihan skala analitik adalah proses penilaian lebih detail karena skala analitik terdiri dari beberapa kriteria penilaian. Jadi, penilai memberikan beberapa nilai untuk tiap kriteria penilaian kemudian menjumlahkan angka-angka tersebut untuk mendapatkan nilai akhir. Dengan demikian, skala analitik sangat bagus dalam memberikan balikan kepada murid dan pengajar. Hanya, kekurangan skala analitik adalah memakan waktu yang lama dalam proses penilaian.

Oleh karena itu, jika lembaga pengajaran bahasa Indonesia ingin memberikan nilai terhadap tulisan siswa dan juga ingin melihat di mana letak kekurangan dari setiap siswa, skala analitik sangat cocok untuk digunakan. Para murid pun dapat memacu diri untuk mengejar kekurangan mereka jika mereka dapat mengetahui letak kekurangannya.

Bentuk skala analitik, seperti yang telah disebutkan oleh Weigle (2002), Moskal (2001), dan Coombe (2010), memisahkan komponen yang dapat diamati dari performa yang dihasilkan oleh siswa untuk kemudian dinilai secara terperinci dan terpisah. Suatu penilaian dapat dikatakan memiliki unsur praktis jika penilaian tersebut tidak mahal, cepat, mudah dilakukan, dan memiliki prosedur penilaian yang spesifik dan efisien (Brown, 2004).

Andrade (1997) dan Underhill (1987) sependapat bahwa skema penilaian yang baik tidak perlu memasukkan terlalu banyak tingkatan skala. Mereka juga menyatakan bahwa skala penilaian tersebut sebaiknya berjumlah genap. Jumlah skala yang genap ini karena jumlah yang genap dapat menghilangkan kemungkinan dari penilai untuk tidak berpihak pada suatu deskripsi tertentu.

Common European Framework (*Council of Europe*, 2001) menyarankan empat sampai lima kriteria sudah cukup untuk digunakan dalam satu buah skema penilaian. Jika menggunakan sampai tujuh kriteria, itu sudah melebihi batas dan dapat menyebabkan penilai justru tidak maksimal dalam memberikan penilaian.

Namun, pemakaian lima sampai enam kriteria masih dapat ditolerir oleh Common European Framework.

Pengajar juga terkadang tidak konsisten dalam menilai sebuah tulisan (*intra-rater reliability*). Hal ini bisa disebabkan ketidakpahaman si pengajar akan skema penilaian yang ia gunakan. Karena itu, keandalan penilaian ini bisa terancam. Selain itu, keandalan antarpilai (*inter-raters reliability*) juga perlu diperhatikan. Sebuah tulisan dapat memperoleh nilai yang berbeda apabila diperiksa oleh penilai yang berbeda walaupun menggunakan skala penilaian yang sama.

Menurut Weigle (2002, hlm. 72), "*Raters bring their own backgrounds, experiences, and values to the assessment of writing.*" Oleh karena itu, dibutuhkan skema penilaian serta pelatihan yang intensif agar para penilai memiliki pandangan dan cara menilai yang tidak jauh berbeda.

McNamara (1996, hlm. 123) juga menyebutkan empat faktor yang menyebabkan perbedaan di antara para pengajar dalam memberikan nilai.

1. *Two raters may simply differ in their overall leniency.*
2. *Raters may display particular patterns of harshness or leniency in relation to only one group of candidates, not others, or in relation to particular tasks, not others.*
3. *Raters may differ from each other in the way they interpret the rating scale they are using.*
4. *Raters may differ in terms of their consistency (or inconsistency); that is, the extent of the random error associated with their ratings.*

Dari penjelasan di atas, peningkatan kualitas skema penilaian diperlukan agar sebuah tulisan dapat dinilai oleh beberapa penilai dengan nilai yang tidak akan jauh berbeda.

Pada kelas menulis di lembaga pengajaran bahasa Indonesia untuk penutur asing, setiap tingkat memiliki kemahiran tertentu yang harus dicapai oleh siswa. Tujuan pembelajaran pada setiap tingkat itu berbeda-beda. Di lembaga pengajaran bahasa Indonesia untuk penutur asing, kemahiran menulis diberikan dengan tingkat kesulitan yang bergradasi, dan siswa harus menguasai kemahiran menulis yang diberikan pada tingkat dasar untuk dapat mengerti kemahiran menulis yang diberikan pada tingkat madya. Oleh karena itu, untuk dapat melanjutkan ke

tingkat berikutnya, seorang siswa harus dapat memiliki kemahiran yang sedang dipelajari pada tingkatnya.

Pada kelas menulis tingkat dasar di lembaga pengajaran bahasa Indonesia untuk penutur asing, kemahiran yang harus dicapai siswa adalah mampu membuat tulisan yang bersifat narasi (cerita). Pada akhir semester, siswa diharapkan sudah menghasilkan sebuah tulisan narasi. Pembuatan tulisan narasi ini melalui proses pembimbingan dengan dosen yang mengajar mata kuliah kemahiran menulis selama satu semester. Siswa diberi beberapa pilihan topik untuk membuat tulisan narasi, misalnya pengalaman ketika sakit, pengalaman jalan-jalan, menulis surat, dan sebagainya.

Pada tingkat madya, tujuan utama pembelajaran kelas menulis adalah siswa mampu membuat tulisan deskripsi. Sama halnya dengan kelas menulis tingkat dasar, pada akhir semester siswa harus membuat sebuah tulisan deskripsi. Proses pembuatan tulisan deskripsi itu juga melalui bimbingan dengan dosen kelas menulis itu. Selain membuat tugas akhir berupa tulisan deskripsi itu, pada ujian akhir semester siswa juga diharuskan membuat tulisan deskripsi berdasarkan topik yang diberikan. Topik yang diberikan ada lebih dari satu, dan siswa hanya memilih satu buah topik. Pada tingkat madya ini, ada empat jenis tulisan deskripsi yang harus dikuasai oleh siswa, yakni tulisan deskripsi diri/orang, deskripsi tempat, deskripsi kegiatan, dan deskripsi perjalanan.

Pada tingkat mahir, kelas menulis dibagi menjadi dua, yakni kelas menulis populer dan kelas menulis ilmiah. Tujuan pembelajaran dari kelas menulis populer adalah siswa mampu membuat lima jenis tulisan populer, yakni advertorial, ulasan film, profil, artikel informatif, dan cerita pendek. Sementara itu, pada kelas menulis ilmiah, siswa harus mampu membuat sebuah tulisan penelitian. Tulisan ilmiah itu minimal berjumlah 30 halaman dengan bagian isi sekitar 15 halaman.

Pada saat ujian tengah semester dan ujian akhir semester, hasil dari ujian tentu akan dinilai oleh pengajar untuk menentukan nilai yang didapat oleh para murid. Namun, setiap lembaga pengajaran bahasa Indonesia untuk penutur asing memiliki skema penilaian untuk menilai tulisan dengan standar dan bentuk yang

berbeda-beda. Mungkin cara pembuatan skema penilaian itu juga menggunakan langkah-langkah pembuatan skema penilaian yang berbeda pula.

Menurut Mertler (2001), sebelum merancang suatu skema penilaian, kita harus menentukan apa yang akan dinilai dan bagaimana cara penilaiannya. Tugas tersulit dalam penyusunan skema penilaian ini adalah ketika mengubahnya ke dalam bentuk nilai. Hal lain yang perlu diperhatikan dari pernyataan Mertler adalah perubahan suatu penilaian menjadi nilai merupakan proses kreatif. Oleh karena itu, pengajar yang bersangkutan harus kreatif dalam menentukan sistem konversi itu. Sistem konversi itu tentunya harus disesuaikan dengan kebutuhan dan tujuan pembelajaran itu sendiri. Sebagai tambahan, Barbara Moskal (2000) menyebutkan bahwa penyusunan suatu rubrik/skala dapat didasarkan pada tujuan akhir dari tugas atau kegiatan yang diberikan kepada siswa.

Oleh karena itu, berdasarkan data di atas, penulis berniat untuk membuat sebuah skema penilaian yang dapat digunakan oleh para pengajar kelas menulis tingkat madya di lembaga pengajaran bahasa Indonesia untuk penutur asing agar nilai yang diberikan terhadap tugas menulis siswa memiliki tingkat keandalan tinggi. Penulis mencoba untuk membuat sebuah skema penilaian mengikuti langkah-langkah pembuatan skema penilaian yang baik agar hasil skema penilaian ini dapat digunakan di seluruh lembaga pengajaran bahasa Indonesia untuk penutur asing. Skema penilaian yang akan dibuat ini mengikuti langkah-langkah pembuatan skema penilaian yang berlandaskan berbagai teori, seperti teori pembuatan skema penilaian oleh Mertler (2001) dan teori penentuan skala penilaian oleh Common European Framework (2001).

Selain itu, penulis juga ingin mencoba untuk mengadaptasi skema penilaian yang sering digunakan dalam pembelajaran bahasa Inggris sebagai bahasa asing. Penulis ingin melihat apakah skema penilaian yang sudah banyak digunakan itu juga dapat digunakan secara tepat untuk menilai tulisan bahasa Indonesia yang dibuat oleh penutur asing.

Alasan lain penulis ingin melakukan penelitian dalam bidang ini adalah karena sejauh pengamatan penulis di Perpustakaan Universitas Indonesia, belum ada peneliti yang membahas keterampilan menulis bahasa Indonesia oleh penutur asing. Oleh karena itu, penulis berharap hasil dari penelitian ini bisa membuka

jalan bagi penelitian lanjutan mengenai bidang ini dan bidang lain yang masih berkaitan, terutama dalam hal pemakaian skema penilaian untuk menilai tulisan.

1.2 Rumusan Pertanyaan Penelitian

Sebuah skema penilaian harus membantu para penilai dalam memberikan nilai secara mudah dan andal. Oleh karena itu, sebuah skema penilaian harus memiliki komponen-komponen penilaian dengan penjelasan terperinci agar para penilai lebih mudah untuk menilai sebuah tulisan. Selain itu, sebuah skema penilaian diharapkan membantu para penilai agar memiliki keseragaman pandangan sehingga apabila sebuah tulisan dinilai oleh penilai yang berbeda, hasilnya tidak akan jauh berbeda.

Berdasarkan pengamatan penulis, setiap lembaga pengajaran bahasa Indonesia untuk penutur asing masing-masing di Indonesia memiliki skema penilaian sendiri dengan bentuk dan jenis skala yang berbeda-beda. Penulis ingin menawarkan skema penilaian yang dapat digunakan di seluruh lembaga pengajaran bahasa Indonesia untuk penutur asing. Penelitian ini juga berusaha menghasilkan sebuah skema penilaian yang dapat menghasilkan penilaian yang andal. Pertanyaan penelitiannya adalah skema penilaian seperti apakah yang akan membantu para penilai tes menulis di lembaga pengajaran bahasa Indonesia untuk penutur asing memberikan penilaian dengan mudah dan andal.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, penelitian ini bertujuan membuat skema penilaian yang dapat digunakan untuk menilai tugas menulis siswa di kelas menulis tingkat madya lembaga pengajaran bahasa Indonesia untuk penutur asing dengan mudah dan andal. Skema penilaian yang akan ditawarkan adalah skema penilaian yang memberi kemudahan bagi pengajar untuk memberi nilai secara objektif dengan tingkat keandalan yang tinggi.

1.4 Ruang Lingkup Penelitian

Penelitian ini dibatasi hanya di salah satu lembaga pengajaran bahasa Indonesia untuk penutur asing, yaitu pada kelas menulis. Di lembaga bahasa tersebut ada tiga tingkat, yaitu tingkat dasar, tingkat madya, dan tingkat mahir. Di

setiap tingkatan itu tentu memiliki tujuan pembelajaran dan penekanan pembelajaran yang berbeda-beda. Skema penilaian yang digunakan pun berbeda-beda, terutama pembobotan nilainya.

Oleh karena itu, penelitian ini akan dibatasi pada perancangan skema penilaian untuk kelas menulis tingkat madya karena pada tingkat ini siswa mendapat pelajaran menulis yang belum terlalu sulit sehingga skema penilaian yang digunakan pun bisa diterapkan untuk semua jenis tulisan pada tahap ini. Sementara itu, di tingkat akhir, kelas menulis sudah terbagi menjadi dua, yaitu kelas menulis karya populer dan kelas menulis karya ilmiah. Karena itu, skema penilaian yang digunakan untuk tiap kelas itu pun berbeda. Sementara itu, alasan penulis tidak meneliti di tingkat dasar karena pada tingkat ini siswa belum mendapat tugas menulis yang terlalu sulit mengingat kosakata mereka masih terbatas dan belum menggunakan skema penilaian untuk menilai tugas siswa.

Data yang akan dianalisis dalam penelitian ini diambil dari hasil penilaian tulisan ujian tengah semester dan ujian akhir semester enam kelas menulis tingkat madya menggunakan skema penilaian yang dibuat. Data itu kemudian dianalisis untuk menguji skema penilaian yang telah dibuat oleh peneliti.

1.5 Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat memberi manfaat secara teoretis dan secara praktis. Secara teoretis, hasil penelitian ini diharapkan dapat memberi sumbangan penelitian bagi bidang pembelajaran bahasa, terutama mengenai masalah evaluasi kemampuan menulis. Sementara itu, manfaat praktis penelitian ini adalah hasil penelitian ini diharapkan dapat menjadi masukan bagi lembaga bahasa tersebut dalam menggunakan skema penilaian untuk menilai tulisan mahasiswa. Selain itu, penelitian ini juga diharapkan dapat memberikan gambaran mengenai cara-cara pembuatan skema penilaian yang baik untuk menilai tulisan.

BAB 2

TINJAUAN PUSTAKA

Pada bab kedua ini akan diuraikan teori dan pendapat para ahli yang akan dijadikan landasan berpikir dalam mengerjakan penelitian ini. Pertama-tama peneliti akan menguraikan definisi mengenai kemahiran menulis. Selanjutnya, diuraikan mengenai tes dan penilaian, jenis-jenis tes menulis, prosedur penilaian, serta prinsip kesahihan, keandalan, dan kepraktisan.

2.1 Penelitian Terdahulu

Dari penelusuran literatur di perpustakaan dan melalui Internet, ditemukan bahwa penelitian mengenai penilaian, baik untuk kemahiran menulis maupun kemahiran berbicara, serta penelitian mengenai skema penilaian, sudah banyak dilakukan, baik di Indonesia maupun di luar negeri. Para peneliti yang telah melakukan penelitian itu antara lain Alvin Taufik (2009), Yuji Nakamura (2004), Wang Yun (2003), Heidi Goodrich Andrade (2003), Rob Schoonen, Margaretha Vergeer, dan Mindert Eiting (1997), Freedman (1979), Carr (2000), Weir (1990), Bauer (1981), Coombe (2010), dan Al-Fallay (2000).

Alvin Taufik (2009) dari Program Studi Ilmu Linguistik di Fakultas Ilmu Pengetahuan dan Budaya Universitas Indonesia menulis tesis berjudul *Penyusunan Skema Penilaian untuk Kelas Talking English Tingkat Dasar di ILP*. Penelitian itu membicarakan mengenai skema penilaian performa untuk kelas *Talking English* tingkat dasar di ILP. Masalah yang menjadi dasar penelitian ini adalah kelas *Talking English* itu memerlukan sistem penilaian yang memiliki kriteria lebih terperinci agar balikan yang dihasilkan akan bermanfaat bagi siswa di kelas itu. Selain itu, masalah utama dalam kelas *Talking English* itu adalah penilaian yang diberikan menjadi sangat tidak berarti karena tidak menjadi dasar kelulusan dalam kelas. Siswa dapat melanjutkan studi jika mereka aktif dalam kelas yang diikuti. Jadi, dalam kelas *Talking English* tidak ada pengujian yang bersifat sumatif. Namun, pengajar di kelas ini tetap harus memberikan penilaian kepada siswa. Penilaian di dalam kelas ini diberikan dalam bentuk skema penilaian analitik dengan kriteria-kriteria yang juga analitis.

Penelitian yang dilakukan oleh Alvin Taufik itu bertujuan memberi kontribusi kepada penilaian kelas Talking English di ILP melalui rancangan skema penilaian yang dapat digunakan pada setiap kegiatan tatap muka. Pada intinya, penelitian itu bertujuan untuk memberi kemudahan bagi para pengajar kelas Talking English agar mampu memberikan penilaian yang lebih objektif. Dengan demikian, pengajar atau penilai dapat memberikan penilaian yang memiliki keandalan tinggi.

Perancangan skema penilaian untuk kelas Talking English di ILP ini dibuat dengan mempertimbangkan unsur penyusunan alat penilaian yang baik, yaitu praktis, andal, dan sah. Berlandaskan pertimbangan tersebut, skema penilaian dibuat dengan menggunakan langkah penyusunan yang dikemukakan oleh Mertler (2001). Langkah itu terdiri atas penentuan tujuan pembuatan skema penilaian, penentuan objek penilaian, penentuan bentuk skema, penentuan kriteria, skala dan deskripsi naratif, percobaan skema, dan diakhiri dengan penyempurnaan skema penilaian. Sementara, penentuan bentuk skema penilaian yang akan digunakan mengikuti penentuan penyusunan skema penilaian yang digagas oleh Moskal (2000).

Temuan yang didapat dari hasil penelitian itu adalah pengajar di kelas percakapan di Indonesia tidak terlalu menganggap penting penguasaan tata bahasa dalam percakapan. Hal itu terbukti dari rendahnya simpangan baku untuk kategori tata bahasa dari hasil kuesioner yang dilakukan oleh Alvin Taufik. Selain itu, dari hasil analisis terhadap skema penilaian yang digunakan di kelas Talking English menghasilkan kesimpulan bahwa skema penilaian berbasis kompetensi memiliki keandalan, kesahihan, dan kepraktisan yang tinggi bahkan bagi para pengajar yang tidak mengetahui arti dari kompetensi.

Selain itu, penulis juga menemukan penelitian lain mengenai metode skema penilaian. Namun, penelitian ini dilakukan di luar negeri, yaitu di Tokyo, Jepang, pada tahun 2004. Penelitian itu dilakukan oleh Yuji Nakamura dengan judul penelitian *A Comparison of Holistic and Analytic Scoring Methods in the Assessment of Writing*.

Hal yang menjadi latar belakang penelitian itu adalah adanya keunggulan dan kekurangan yang dimiliki oleh skala holistik dan skala analitik. Keuntungan

menggunakan skala holistik untuk menilai tulisan adalah lebih cepat proses penilaiannya karena penilai hanya perlu memberikan satu nilai tunggal. Sementara itu kekurangan dari skala holistik adalah tiap pengajar memiliki fokus penilaian yang berbeda-beda ketika menilai sebuah tulisan. Di lain pihak, keunggulan dari skala analitik adalah tiap penilai memberikan nilai kepada setiap unsur yang membentuk sebuah tulisan. Akan tetapi, kekurangan utama dari penggunaan skala analitik adalah memakan waktu yang cukup lama dalam proses penilaian.

Penelitian yang dilakukan oleh Yuji Nakamura bertujuan menentukan kelemahan dan kekuatan dari metode *holistic scoring* dan *analytic scoring* dengan menggunakan metode dari Weigle yang merupakan adaptasi dari pemikiran Bachman dan Palmer yang memiliki enam kategori kegunaan tes. Selain itu, penelitian ini juga mencoba untuk menentukan cara penilaian mana yang lebih baik dalam menilai siswa, apakah menggunakan skala holistik atau menggunakan skala analitik.

Metode yang digunakan adalah dengan menilai 90 tulisan mahasiswa oleh tiga orang penilai menggunakan skala holistik dan skala analitik. Pada skala analitik yang digunakan ada lima kriteria penilaian, yaitu tata bahasa, kosakata, organisasi, orisinalitas, dan kepaduan. Penilaian dengan kedua skala penilaian dilakukan pada hari yang berbeda dengan jeda waktu tiga minggu. Tulisan yang dinilai oleh ketiga penilai diberikan skala 1-4 (1=buruk, 4=bagus). Analisis data dilakukan menggunakan model FACETS sehingga ketiga *facets* (murid, penilai, dan item evaluasi) dapat ditunjukkan pada satu kontinum.

Semua mahasiswa adalah lulusan baru dan berusia antara 18 sampai 20 tahun, serta separuh lebih berjenis kelamin pria. Para penilai diberi pelatihan menggunakan 10 contoh tulisan yang dikumpulkan dari mahasiswa universitas yang sama. Topik dari tulisan yang akan dinilai adalah mengenai proposal untuk mengisi liburan selama lima hari.

Ada beberapa kesimpulan yang didapat dari penelitian ini. Pertama, untuk alasan kepraktisan dan ekonomis, skala holistik dapat digunakan. Namun, untuk menghindari penilaian yang subjektif, skala analitik lebih direkomendasikan untuk digunakan. Kedua, cara penilaian yang baik adalah dengan menggunakan beberapa penilai dengan beberapa kriteria penilaian. Cara terbaik kedua adalah

dengan menggunakan satu kriteria penilaian, tetapi dinilai oleh beberapa penilai. Cara yang terbaik ketiga adalah dengan satu penilai menggunakan beberapa kriteria penilaian. Terakhir, penelitian ini menyimpulkan bahwa akan sangat berisiko apabila pengajar kelas menilai tulisan siswa menggunakan skala holistik.

Penelitian mengenai penilaian juga penulis temukan dari beberapa jurnal. Salah satunya adalah yang dilakukan oleh Wang Yun (2003) dari negeri Cina dengan judul penelitian *How Raters and Writers Perceptions of a Topic Affect the Scoring of Compositions*. Penelitian ini bertujuan menemukan pengaruh persepsi penulis dan penilai dari topik yang diberikan terhadap nilai komposisi pelajar. Dengan kata lain, apakah siswa mendapat nilai yang bagus jika si penilai setuju dengan apa yang ditulis olehnya? Wang Yun mengajukan beberapa data kuantitatif dengan tujuan menolak *null hypothesis* yang menyatakan bahwa tidak ada hubungan antara sikap atau pandangan si penilai dan nilai tulisan siswa.

Subjek penelitian itu adalah 53 siswa dengan latar belakang pendidikan yang sama. Semua siswa itu adalah mahasiswa senior dari Universitas Harbin Engineering, Republik Rakyat Cina. Metode penelitiannya adalah dengan meminta semua siswa menulis tiga buah komposisi, masing-masing 120 kata, berdasarkan tiga pertanyaan yang diajukan, yaitu (1) Apakah masih bisa ada peningkatan di dalam sistem universitas di Cina? (2) Apakah yang paling diperlukan di masyarakat: kompetisi atau kerja sama? dan (3) Bagaimana pendapat Anda mengenai pemecatan karyawan di Cina?. Kemudian, hasil tulisan tersebut diberikan kepada para penilai. Penilai untuk hasil tulisan siswa ada tiga orang yang merupakan dosen TEFL yang biasa mengajar komposisi atau esai di universitas. Setelah itu, data dianalisis menggunakan *two-way analysis of variance* (ANOVA). Kesimpulan dari analisis menggunakan ANOVA itu adalah tidak ada pengaruh yang signifikan antara sikap atau pandangan si penilai dengan nilai dari tulisan siswa.

Penelitian lain mengenai rubrik atau skema penilaian didapatkan dari *The Journal of Educational Research* vol. 97 tahun 2003 yang berjudul *Role of Rubric-Referenced Self-Assessment in Learning to Write*. Penelitian itu dilakukan oleh Heidi Goodrich Andrade dari Universitas Albany dan Beth A. Boulay dari Abt Associates, Inc., Cambridge, Massachusetts. Mereka meneliti pengaruh *self-*

assessment pada esai mahasiswa tingkat ke-7 dan ke-8. Tiap mahasiswa menulis dua esai, yaitu *historical fiction* dan *response to literature*.

Semua siswa menerima rubrik instruksi mengenai kriteria dan kualitas nilai dari esai yang akan dikerjakan. Siswa dalam grup mendapat *treatment* dalam dua *formal self-assessment lessons*, yakni untuk mengecek draf tulisan mereka dengan menggunakan rubrik yang diberikan. Peneliti menggunakan *multiple linear regression* untuk menentukan hubungan antara nilai esai, *treatment*, dan satu set *control predictors*.

Hasil analisis dari esai *historical fiction* menunjukkan bahwa ada hubungan yang positif antara *treatment* yang diberikan dengan nilai para siswa perempuan. Namun, tidak ada hubungan yang signifikan antara *treatment* yang diberikan dengan nilai para siswa laki-laki. Sementara itu, untuk esai *response to literature* tidak ada pengaruhnya antara *treatment* yang diberikan dengan nilai para siswa, baik itu siswa perempuan atau laki-laki.

Penelitian lain mengenai penilaian kemampuan menulis yang ditemukan oleh penulis adalah *The Assessment of Writing Ability: Expert Readers Versus Lay Readers* yang ditulis oleh Rob Schoonen, Margaretha Vergeer, dan Mindert Eiting dari Universitas Amsterdam pada tahun 1997. Penelitian ini mengenai keandalan membaca antara pembaca awam dan pembaca berpengalaman dalam menilai tiga buah komponen penulisan. Para pembaca harus menilai komponen “Isi” dan “Penggunaan Bahasa” dari tulisan siswa.

Penelitian itu menunjukkan bahwa pembaca berpengalaman lebih andal dalam menilai komponen “Penggunaan Bahasa”, sedangkan dalam menilai komponen “Isi”, baik pembaca awam maupun pembaca berpengalaman sama-sama andal. Kesimpulannya adalah perbedaan keahlian yang dimiliki oleh para penilai sangat berpengaruh pada hasil penilaian.

Freedman (1979) melakukan penelitian dengan membandingkan antara skala holistik dan skala analitik dalam mengevaluasi tulisan mahasiswa dan penulis profesional. Hasilnya menunjukkan bahwa jika menggunakan skala analitik, nilai untuk tulisan mahasiswa dan tulisan penulis profesional berbeda. Namun, jika menggunakan skala holistik, nilai keduanya sama. Selanjutnya, nilai tulisan para mahasiswa itu rata-rata sama, baik menggunakan skala holistik

maupun dengan skala analitik. Sementara itu, para penulis profesional mendapatkan nilai tulisan yang lebih tinggi jika menggunakan skala analitik daripada jika menggunakan skala holistik.

Pada tahun 2000, Carr menginvestigasi pengaruh dari penggantian skema penilaian analitik ke skema penilaian holistik dalam tes penempatan ESL di sebuah universitas. Hasilnya adalah mengubah skema penilaian ikut mengubah penekanan di keseluruhan tes—yang terdiri dari kemahiran menyimak, membaca, dan menulis. Dengan menggunakan skala holistik, nilai condong ke arah kemampuan produktif daripada kemampuan reseptif, padahal tidak ada satu pun bagian tes yang diubah.

Selain itu, juga ada penelitian dari Weir (1990) yang berdasarkan laporan dari Hartog et al. (1936) dan Cast (1939), menyimpulkan bahwa skema penilaian analitik lebih andal daripada skala holistik, walaupun tidak ada penilai yang mendapatkan pelatihan terlebih dahulu. Bauer (1981) juga menemukan bahwa skala analitik lebih andal daripada skala holistik walaupun skala holistik lebih hemat biaya.

Penelitian berikutnya yang penulis temukan adalah dari Christine Coombe. Penelitian yang dilakukan pada tahun 2010 ini berjudul *Assessing Foreign/Second Language Writing Ability*. Tujuan dari penelitian ini adalah untuk memberikan gambaran kepada para pengajar mengenai masalah-masalah yang sering dihadapi ketika menilai sebuah tulisan. Penelitian ini akan memperlihatkan isu-isu dan solusi dalam lima area utama, yaitu desain tes, penyelenggaraan tes, cara menilai tulisan, memberi balikan kepada siswa, dan pengaruhnya terhadap pendidikan.

Metode yang digunakan oleh Coombe adalah dengan pendekatan praktis terhadap proses penilaian tulisan siswa dalam bahasa kedua atau bahasa asing. Temuan dalam penelitian ini adalah bahwa proses pengajaran dan penilaian berkaitan dengan pencapaian siswa, bukan pencapaian institusi. Implikasi praktis penelitian ini adalah pengajar yang baik akan meluangkan waktu yang banyak untuk memastikan bahwa penilaian yang diberikan sudah andal dan sah.

Penelitian terakhir yang penulis temukan mengenai skema penilaian adalah yang dibuat oleh Ibrahim Al-Fallay. Penelitian yang dilakukan pada 2000 ini berjudul *Examining the Analytic Marking Method: Developing and Using an*

Analytic Scoring Schema. Tujuan dari penelitian ini adalah untuk menjawab pertanyaan apakah mungkin untuk membuat sebuah skema penilaian yang memiliki keandalan intrapenilai dan keandalan antarpemilai yang tinggi. Lalu, fitur atau komponen apa saja yang harus ada di dalam skema penilaian itu? Kemudian, bagaimana pemberian bobot untuk setiap komponen itu?

Subjek penelitian ini adalah 55 lulusan SMU di Arab Saudi yang mendaftar ke Departemen Inggris di Universitas King Saud. Para mahasiswa itu diminta untuk membuat tulisan berdasarkan topik yang diberikan. Mereka diberi waktu 30 menit untuk membuat tulisan dengan jumlah paragraf minimal 12. Tulisan tersebut kemudian dinilai secara analitik oleh tiga pemilai. Penelitian ini menyimpulkan bahwa penilaian analitik dapat digunakan untuk melihat kekurangan dan kelebihan murid.

Dari beberapa penelitian terdahulu yang telah diuraikan, terlihat bahwa skala analitik lebih banyak digunakan dan memiliki keunggulan dibandingkan skala holistik dalam menilai tulisan. Hal tersebut menjadi salah satu pertimbangan bagi penulis untuk menyusun skema penilaian analitik dibanding skema penilaian holistik. Dari penelitian terdahulu itu juga penulis dapat melihat langkah-langkah pembuatan skema penilaian yang baik, dan penulis mencoba menerapkan langkah-langkah pembuatan skema penilaian itu untuk membuat skema penelitian dalam penelitian ini.

Selain itu, dari beberapa penelitian yang telah dijabarkan di atas, terlihat bahwa penelitian mengenai pembuatan skema penilaian untuk menilai tulisan siswa di kelas menulis tingkat madya di lembaga pengajaran bahasa Indonesia untuk penutur asing belum pernah ada yang melakukan. Oleh karena itu, penulis ingin membuat skema penilaian yang dapat digunakan di lembaga pengajaran bahasa Indonesia untuk penutur asing. Penulis juga berharap hasil penelitian ini dapat menjadi batu loncatan bagi peneliti lain untuk melakukan penelitian lanjutan.

2.2 Kemahiran Menulis

Blanton (1995, dlm Raimes, 1998) membandingkan definisi menulis seperti orang buta yang memegang seekor gajah. Persepsi yang dihasilkannya bisa bermacam-macam. Pengajar bisa mengartikan menulis hanya sebuah teks yang

mencontoh sebuah model; pengajar bisa menganggap menulis sebagai representasi dari realitas yang dialami si penulis; atau menulis sebagai permintaan dari para pembaca. Kemahiran menulis memang tidak memiliki satu definisi yang cukup umum sebab tulisan pun memiliki variasi, dan tiap orang mempunyai definisi yang berbeda akan tiap jenis tulisan. Bagi pemelajar bahasa kedua, belajar menulis bisa memiliki arti yang luas, mulai dari menulis surat sampai menulis skripsi. Brown (2004) mengklasifikasi tulisan menjadi tiga genre, yaitu *academic writing*, *job-related writing*, dan *personal writing*.

Untuk membuat sebuah tulisan yang bagus dalam bahasa pertama, tidak semua orang pandai melakukannya, apalagi jika harus membuat tulisan dalam bahasa kedua. Karena itu, bisa dikatakan kemahiran menulis bukanlah hal yang mudah untuk dikuasai dalam pembelajaran bahasa kedua. Seorang pemelajar bahasa kedua yang sudah mahir dalam berbicara menggunakan bahasa kedua belum tentu mahir dalam menuliskannya, sebab dalam membuat sebuah tulisan ada banyak faktor bahasa yang harus dipertimbangkan, seperti ejaan, tata bahasa, dan tanda baca.

Weigle (2002) menyatakan bahwa kurangnya pengetahuan linguistik dalam bahasa kedua membuat proses penulisan dalam bahasa kedua terhambat karena siswa lebih fokus pada bahasa daripada isi. Karena itu, hasil tulisannya pun bisa jauh dari apa yang diinginkan semula. Selain itu, pemelajar bahasa kedua juga memiliki keterbatasan dalam faktor sosial dan budaya bahasa target. Akibatnya adalah mereka kurang mampu menggunakan bentuk-bentuk bahasa yang cocok dengan konteks lingkungan atau latar belakang para pembaca.

Dalam kemahiran menulis, ada beberapa kemampuan yang perlu dipelajari dalam rangka menguasai kemahiran menulis ini. Misalnya, dalam penggunaan bahasa, siswa mampu untuk menulis kalimat secara baik dan benar. Dalam pengembangan gagasan, siswa mampu berpikir secara kreatif dan mampu mengembangkan pemikiran, serta memisahkan informasi yang tidak termasuk dalam ide penulisan. Dalam *judgement skill*, siswa mampu membuat tulisan dengan laras bahasa yang cocok dengan pembacanya. Kemampuan dalam kemahiran menulis yang sangat penting untuk dikuasai, terutama untuk tingkat dasar, adalah tanda baca dan ejaan. Yang lebih penting lagi adalah kemampuan

untuk menentukan jenis tulisan, yakni kemampuan untuk menulis bagi pembaca yang spesifik serta menggunakan laras kalimat yang sesuai. Kemampuan tersebut sangat penting untuk dikuasai, baik oleh penutur jati maupun pelajar bahasa kedua.

Penilaian tulisan pun tidak mudah, sebab pengajar atau penilai harus menentukan secara spesifik hal apa yang ingin dinilai, apakah ejaannya, tata bahasanya, konstruksi paragraf, atau pengembangan gagasannya. Tiap hal tersebut juga dapat dilakukan dengan berbagai macam jenis tes. Menurut Blok dan Glopper (1992), ada tiga alasan kenapa harus dilakukan penilaian kemahiran menulis, yakni penilaian adalah cara yang efisien untuk mengukur pencapaian kemahiran menulis, penilaian dapat berpengaruh dalam perkembangan kemampuan pelajar serta kurikulum pembelajaran, dan penilaian sangat berguna bagi pengembangan prosedur pengukuran yang lebih andal dan sah.

Tugas yang mengharuskan siswa membuat sebuah tulisan pada saat dilakukan penilaian disebut juga sebagai tes menulis langsung (*direct test*). Tes langsung adalah metode yang paling sering digunakan untuk menguji kemampuan menulis, baik dalam konteks bahasa pertama maupun bahasa kedua (Weigle, 2002).

Breland (1983, dalam Blok dan Glopper, 1992) membagi dua pendekatan untuk mengukur kemahiran menulis, yaitu penilaian langsung (*direct assessment*) dan penilaian tak langsung (*indirect assessment*). Menurut Breland, dalam penilaian langsung, contoh dari tulisan siswa dinilai oleh satu atau lebih dari satu penilai. Sementara itu, penilaian tak langsung tidak membutuhkan contoh tulisan siswa. Penilaian tak langsung mengukur kemahiran siswa dari performa tugas yang lain.

Coombe (2010) menambahkan bahwa tes langsung mengharuskan siswa memerhatikan mengenai isi tulisan, organisasi ide, serta menggunakan kosakata, tata bahasa, dan sintaksis yang tepat. Tes langsung ini menggabungkan semua elemen dari kemampuan menulis. Shaw dan Weir (2007, hlm. 9) juga menambahkan bahwa “*By a direct test, we mean one which tests writing through involving candidates in the actual construction of text in contrast to ‘indirect’ or*

'objective' tests of writing which principally focus on knowledge of microlinguistics elements of writing.'

Hamp-Lyons (1991a, hlm. 5-6) menyarankan bahwa tes menulis langsung minimal harus memiliki unsur-unsur di bawah ini:

1. Sampel tulisan siswa minimal harus berjumlah 100 kata dalam satu tulisan.
2. Penulis diberikan keleluasaan untuk merespons instruksi yang diberikan walaupun sudah diberikan masukan materi yang spesifik.
3. Setiap tulisan dibaca minimal oleh satu atau dua orang penilai.
4. Penilaian yang diberikan harus mengikuti standar yang telah ditetapkan.
5. Penilaian harus dinyatakan dalam bentuk angka, bukan disampaikan secara lisan atau berupa komentar.

Akan tetapi, penilaian kemahiran menulis secara langsung juga menimbulkan satu masalah lain, yakni keandalan. Problem untuk penilaian kemahiran menulis dalam jumlah banyak adalah keandalan. Menurut Quellmalz (1980, dalam Weigle, 2002), *"Raters tend to drift away from each other and from the criteria they used at the beginning."* Salah satu faktor yang menyebabkan ketidakandalan itu adalah kelelahan. Oleh karena itu, pada lembaga pengajaran bahasa biasanya membatasi jumlah murid per kelasnya.

Pada tulisan yang menjadi sampel penelitian ini, penulis menggunakan sampel dari kelas menulis tingkat madya. Pada tingkat madya, jenis tulisan yang dihasilkan adalah sebuah tulisan deskripsi. Namun, ada sedikit perbedaan dari hasil tulisan yang dihasilkan pada ujian tengah semester dan ujian akhir semester. Pada ujian tengah semester, para siswa hanya diminta untuk membuat tulisan deskripsi dengan panjang satu paragraf, sedangkan pada ujian akhir semester, para siswa diminta untuk membuat sebuah tulisan deskripsi yang cukup panjang dengan jumlah paragraf minimal lima.

2.3 Tes dan Penilaian

Banyak orang menganggap bahwa pengertian tes (*test*) dan penilaian (*assessment*) sama. Namun, pendapat itu keliru. Pengertian penilaian yang saya gunakan dalam penelitian ini merujuk pada pendapat Brown (2004) yang menyatakan bahwa penilaian merupakan suatu proses berkelanjutan dalam

mengukur performa siswa. Sementara itu, tes adalah sebuah alat untuk mengukur performa, kemampuan, dan pengetahuan seseorang dalam suatu ranah tertentu. Hughes (2003) juga menambahkan bahwa tes merupakan salah satu bentuk dari penilaian. Sementara itu, menurut Rahayu S. Hidayat (1990, hlm. 47), “Pengetesan adalah proses atau tindakan memeriksa kemampuan, dan dalam bidang pengajaran kemampuan yang diperiksa adalah hasil pembelajaran.” Dalam pengajaran bahasa, pengetesan lazim dilakukan untuk mengukur pengetahuan bahasa, seperti lafal, kosakata, dan tata bahasa, serta keempat keterampilan bahasa, yaitu menyimak, berbicara, membaca, dan menulis.

Lebih lanjut Brown mengungkapkan bahwa tes dapat mengacu pada suatu norma (*norm-referenced test*) atau pada suatu standar kriteria (*criterion-referenced test*). Pada tes yang mengacu ke norma, nilai tes dihubungkan dengan serangkaian perhitungan statistik, seperti nilai rata-rata (*mean*), median, atau simpang baku. Tes TOEFL merupakan salah satu contoh jenis tes tersebut. Sementara itu, tes yang mengacu pada suatu kriteria, tes tersebut dirancang untuk memberi balikan atau masukan bagi siswa yang mengikuti tes itu. Tes yang dilakukan di dalam kelas merupakan contoh dari jenis tes tersebut. Untuk tes jenis ini, peran pengajar dalam menyusun tes sangat penting karena tes itu nantinya digunakan sebagai alat pemberi masukan bagi siswa.

Selain kedua pembagian besar tes di atas, Brown juga membagi jenis tes berdasarkan tujuannya, yaitu tes bakat bahasa (*aptitude test*), tes kemahiran (*proficiency test*), tes penempatan (*placement test*), tes diagnostik (*diagnostic test*), dan tes keberhasilan (*achievement test*). Tes kemampuan secara khusus bertujuan mengetahui kesuksesan seseorang dalam mempelajari bahasa asing. Tes pemrakira dilakukan untuk menguji kompetensi global seseorang dalam mempelajari bahasa asing. Tes semacam ini sulit untuk dibuat, seperti tes TOEFL. Tes penempatan dilakukan untuk menempatkan siswa di tingkat yang tepat ketika mereka memutuskan untuk mempelajari suatu bahasa asing. Tes penempatan mempunyai banyak variasi penilaian. Tes semacam ini dapat digunakan untuk mengukur pemahaman, produksi kebahasaan siswa, dan berbagai jenis kompetensi. Tes diagnostik dilakukan jika pengajar ingin mengetahui kekurangan atau kelebihan siswa dalam mempelajari suatu bahasa. Terdapat perbedaan yang

sangat besar antara tes diagnostik dengan tes pencapaian. Pada tes pencapaian yang diukur adalah pemahaman siswa terhadap bahasa yang sedang dipelajarinya. Tes pencapaian sangat dianjurkan untuk disusun berdasarkan silabus yang dipakai. Tes pencapaian itu dapat dilakukan untuk mengukur perkembangan kebahasaan siswa (*progress test*) atau pemahaman siswa atas keseluruhan pelajaran dalam kurun waktu tertentu (*final test*).

Selain pembedaan tes itu, Brown juga membagi penilaian atas dua kategori, yaitu penilaian formal dan penilaian informal. Penilaian informal yang dimaksud di sini dapat berupa komentar atas performa siswa dalam melakukan kegiatan dan pemberian saran atau masukan tentang hasil karya mereka. Contoh bentuk penilaian informal adalah penilaian terhadap jurnal berkala atau portofolio yang dibuat siswa selama proses belajar mereka. Sementara itu, penilaian formal dibuat secara sistematis untuk mempermudah pengajar mendapatkan informasi yang bermanfaat, baik bagi mereka maupun bagi siswa sehingga mereka dapat mengetahui performa mereka dengan jelas.

Penilaian juga dapat dibedakan berdasarkan fungsinya, yaitu penilaian sumatif dan penilaian formatif. Penilaian sumatif berfungsi mengukur pemahaman siswa atas suatu pelajaran atau unit instruksi. Penilaian lain yang berdasarkan fungsi adalah penilaian formatif. Brown mengungkapkan bahwa semua penilaian informal merupakan penilaian formatif. Fokus utama dari penilaian semacam ini adalah perkembangan kebahasaan siswa secara berkelanjutan.

Hughes (2003) berpendapat bahwa penilaian dikatakan formatif apabila pengajar menggunakannya untuk mengecek perkembangan siswa mereka, melihat sejauh mana siswa menguasai materi yang telah dipelajari, dan menggunakan informasi tersebut sebagai pertimbangan bahan mengajar mereka selanjutnya. Salah satu tes yang dapat digunakan dalam penilaian formatif adalah kuis atau tes informal. Penilaian sumatif biasanya diadakan pada akhir semester atau tahun untuk mengukur apa yang telah dicapai oleh siswa. Pada penilaian sumatif, jenis tes yang digunakan adalah yang bersifat formal.

Menurut Coombe, ada tiga hal yang perlu diperhatikan dalam penilaian formal dan sumatif, yakni siapa yang akan menilai, jenis skema penilaian yang akan digunakan, dan prosedur untuk menilai tulisan. Untuk menilai tulisan, ada

dua pilihan yang dapat digunakan, yaitu apakah si pengajar kelas tersebut yang menjadi penilai atau menggunakan beberapa penilai.

Penggunaan pengajar sebagai penilai tulisan menimbulkan pro dan kontra di antara para ahli pengajaran. Para ahli yang kontra dengan pengajar yang juga menjadi penilai tulisan berpendapat bahwa hal itu dapat menimbulkan kemungkinan subjektifitas si pengajar terhadap murid tertentu. Seperti yang Coombe (2010, hlm. 183) ungkapkan, “*There is possibility that teachers might show bias either for or against a particular student.*” Sementara itu, para ahli yang pro akan pengajar yang menjadi penilai mengatakan bahwa para pengajar itulah yang paling mengerti kemampuan para siswa. Karena itu, mereka juga harus menjadi penilai tulisan mereka. Coombe berpendapat bahwa akan lebih baik apabila menggunakan metode *double blind marking* karena penilai tidak mengetahui informasi maupun nama dari penulis yang sedang dinilai. Cara yang kedua untuk menilai tulisan adalah dengan menggunakan lebih dari satu penilai. Alasannya adalah nilai yang dihasilkan dari para penilai itu mendekati nilai yang sebenarnya dari kemampuan siswa dibandingkan jika hanya menggunakan satu penilai (Hamp-Lyons, 1990).

2.4 Jenis Tugas Menulis

Brown (2004) membagi performa tugas menulis menjadi empat kategori, yaitu:

1. *Imitative*. Pada kategori ini, pemelajar diharapkan menguasai kemampuan menulis dasar, seperti mengeja dengan benar. Jenis tugas yang biasa diberikan adalah menulis huruf, kata, atau tanda baca. Pada tahap ini, fokus utamanya adalah bentuk, sedangkan konteks dan arti merupakan perhatian nomor dua.
2. *Intensive*. Pada kategori ini siswa diharapkan dapat menggunakan kosakata yang sesuai dengan konteks, kolokasi, idiom, dan tata bahasa yang tepat untuk tingkat kalimat. Konteks dan arti menjadi perhatian utama karena menjadi penentu berterima atau tidaknya kalimat yang dibuat.
3. *Responsive*. Pada tahap ini, tugas yang diberikan mengharuskan siswa menampilkan performa dalam sebuah wacana, merangkai kalimat menjadi sebuah paragraf, dan membuat sebuah hubungan logis yang menghubungkan

dua-tiga paragraf. Jenis tugas yang diberikan antara lain membuat tulisan narasi atau deskripsi, ringkasan, atau interpretasi sebuah grafik.

4. *Extensive*. Pada tahap ini siswa dianggap sudah menguasai semua proses dan strategi penulisan untuk semua jenis tulisan. Siswa sudah dapat membuat esai, makalah, laporan proyek, bahkan tesis. Dalam tahap ini, fokus utama penulisan adalah mencapai tujuan dari tulisan itu, organisasi dan pengembangan ide secara logis, menggunakan contoh-contoh untuk mendukung ide utama, dan menggunakan sintaksis dan leksikal lebih bervariasi. Sementara itu, bentuk-bentuk gramatikal hanya sebatas *proofreading*.

2.4.1 Jenis Tes Tulisan *Imitative*

Melihat penjelasan tulisan *imitative* yang sudah dijelaskan di atas, kita dapat menyimpulkan bahwa ini merupakan tingkat dasar dari kemahiran menulis. Oleh karena itu, tes yang diberikan belum terlalu sulit, masih pada tataran kata. Beberapa jenis untuk tulisan *imitative* adalah:

1. Mengkopi

Pada tes ini, siswa diminta menuliskan kembali atau mengkopi kata yang dijadikan acuan. Kata yang dijadikan acuan itu bisa ditambahkan dengan tanda baca, misalnya *Oh?* Atau *Ya?*

2. Isi rumpang

Pada tugas isi rumpang, siswa diberi sebuah paragraf yang memiliki rumpang di beberapa bagian. Tugas siswa adalah untuk mengisi bagian yang rumpang itu dengan kata-kata yang cocok. Biasanya, pengajar memberikan pilihan kata-kata yang menjadi jawaban soal itu, tapi diletakkan secara acak. Untuk lebih mempersulit, pilihan kata itu ditiadakan. Bagian yang rumpang itu biasanya muncul secara teratur, misalnya setiap empat atau lima kata.

3. Tugas bergambar (*picture-cued task*)

Pada tugas ini, siswa diberikan gambar yang sudah familiar, kemudian siswa diminta untuk menuliskan nama atau gambar apa yang dimaksud itu. Gambar yang diberikan tidak yang menimbulkan ambiguitas di antara para siswa.

4. Melengkapi isian

Pada tes ini siswa diminta melengkapi atau mengisi sebuah formulir, misalnya mengisi formulir biodata. Pada formulir sudah ada kata-kata seperti *nama*, *alamat*, *nomor telepon*, dan lainnya, kemudian siswa diminta untuk mengisi data-data yang diminta tersebut.

5. Mengubah angka dan singkatan menjadi kata-kata

Sebenarnya tes jenis ini jarang sekali digunakan dan kurang autentik, sebab pada bentuk nyata, jarang sekali orang yang menuliskan angka dalam bentuk kata-kata, kecuali dalam beberapa kasus, seperti menulis cek. Namun, mengubah singkatan menjadi bentuk lengkap masih bisa masuk akal dan autentik.

6. Mengeja

Pada tes ini pengajar mendikte sejumlah kata, lalu merangkainya menjadi kalimat. Kemudian, siswa diharuskan menulis kembali apa yang telah diucapkan oleh pengajar berdasarkan apa yang mereka ingat. Penilaian tes ini berdasarkan benar-tidaknya ejaan kata-kata yang dituliskan. Tes mengeja ini sudah sangat tradisional, dan bersinggungan dengan pengetesan kemahiran menyimak.

2.4.2 Jenis Tes Tulisan *Intensive*

Tahap selanjutnya dari kemahiran menulis adalah yang disebut tulisan terkontrol. Pada tahap ini, siswa menampilkan kompetensi menulis mereka dalam tata bahasa, kosakata, dan rangkaian kalimat. Ada empat jenis tes untuk menilai tulisan *intensive*, yaitu:

1. *Dicto-comp*

Pada tes ini, pengajar membaca sebuah paragraf dengan kecepatan normal, biasanya dua sampai tiga kali, kemudian siswa disuruh menulis ulang paragraf tersebut berdasarkan kata-kata yang mereka ingat.

2. Tugas bergambar (*picture-cued task*)

Tugas bergambar ini dibagi lagi menjadi tiga jenis, yakni menulis kalimat pendek (*short sentences*), deskripsi gambar, dan deskripsi gambar berurutan (*picture sequence description*). Pada tugas menulis kalimat pendek, siswa diminta untuk menuliskan aktivitas apa yang sedang dilakukan pada gambar yang diberikan. Biasanya, di bawah gambar yang diberikan ada pertanyaan yang harus dijawab, misalnya pertanyaan “Apa yang dilakukan perempuan itu?”, dan siswa harus menuliskan jawaban berdasarkan pertanyaan dan gambar yang diberikan.

Pada tugas deskripsi gambar, siswa diminta untuk mendeskripsikan sebuah gambar yang lebih kompleks. Jika pada tugas menulis beberapa kalimat pendek hanya diberikan satu gambar aktivitas, misalnya seorang perempuan sedang membaca buku, pada tugas deskripsi gambar ini ada beberapa hal atau aktivitas yang harus dideskripsikan oleh siswa, misalnya ada seorang pria yang duduk di atas sofa sedang menonton TV, lalu di bawah meja ada kucing yang sedang makan, dan sebagainya. Biasanya, untuk tugas jenis ini, ada beberapa hal yang harus digunakan, misalnya harus memakai kata-kata *di atas*, *di bawah*, *di samping*, atau *di sebelah*.

Pada tugas bergambar yang ketiga, yaitu deskripsi gambar berurutan, siswa diminta untuk membuat sebuah cerita berdasarkan gambar-gambar yang diberikan. Gambar-gambar itu disusun secara berurutan sehingga membentuk suatu jalinan cerita. Gambar yang diberikan harus sederhana dan tidak menimbulkan keambiguan.

3. Menyusun kata

Pada tugas ini, siswa diberikan kata-kata yang disusun secara acak. Kemudian, siswa diminta untuk menyusun kata-kata itu menjadi sebuah kalimat yang benar. Tugas seperti ini kurang autentik untuk mengukur performa menulis. Tugas menyusun kata ini lebih melibatkan kemampuan membaca siswa daripada kemampuan menulis.

4. Jawaban singkat dan meneruskan kalimat

Pada tugas ini, siswa diminta untuk menulis jawaban singkat dari pertanyaan yang diberikan, misalnya siswa diminta untuk menulis jawaban dari pertanyaan “Dari mana asalmu?”. Jawaban yang diberikan harus benar secara konteks dan tata bahasa. Dalam tes membuat kalimat, siswa diberikan suatu situasi, misalnya sedang berada di pasar, dan siswa diharuskan untuk menulis kalimat-kalimat yang berhubungan dengan jual-beli atau menawar harga.

2.4.3 Jenis Tes Tulisan *Responsive* dan *Extensive*

Dalam membuat tulisan *responsive* dan *extensive*, tugas yang diberikan bersifat *open-ended*, seperti menulis laporan singkat, esai, atau rangkuman. Tulisan yang dihasilkan bisa berjumlah satu halaman atau lebih.

1. Parafrase

Tujuan dari parafrase adalah agar siswa dapat menuliskan sesuatu dengan kata-kata sendiri untuk menghindari plagiarisme dan untuk mengungkapkan ekspresi yang lebih beragam. Tugas parafrase lebih condong ke penilaian informal dan formatif daripada penilaian formal dan sumatif. Oleh karena itu, dari tugas parafrase ini diharapkan pengajar mendapat *washback* positif untuk kepentingan mengajar ke depannya.

2. Membuat Paragraf

Mengembangkan ide utama dan ide pendukung merupakan salah satu tujuan dari penulis untuk membuat tulisan yang efektif, baik hanya satu-dua paragraf atau tulisan panjang yang sampai lebih dari dua halaman. Pada tulisan yang

dihasilkan oleh siswa, ada beberapa komponen penilaian yang harus diperhatikan oleh penilai, yakni kalimat topik, pengembangan topik dalam paragraf, dan pengembangan ide utama dan ide pendukung antarparagraf.

3. *Test of Written English* (TWE)

Salah satu tes untuk menguji kemahiran menulis bahasa Inggris yang sudah diakui secara internasional adalah *Test of Written English* (TWE). Pada TWE, siswa tidak diberi tahu terlebih dahulu mengenai topik yang harus ditulis. Topik tulisan baru akan diberikan ketika tes dimulai. Siswa diberi waktu selama 30 menit untuk membuat tulisan dari topik yang diberikan.

Senada dengan Brown, Weir (2005) juga menyebutkan beberapa jenis tes untuk kemahiran menulis, yaitu isi rumpang (*gap filling*), *open-ended essay tests*, dan merespons informasi yang diberikan. Pada tes *open-ended essay tests*, siswa diberi satu topik, misalnya “liburan”, kemudian siswa diminta untuk menuliskan pengalamannya selama liburan. Sementara itu, dalam tes memberikan respons terhadap informasi yang diberikan, siswa diminta untuk membuat tulisan menanggapi informasi yang telah diberikan oleh pengajar. Salah satu contohnya adalah siswa diminta untuk menulis surat sebagai tanggapan atas surat sebelumnya. Weir (2005) menjelaskan bahwa tugas dalam kemahiran menulis dapat bersifat referensial (bertujuan memberi informasi), konatif (bertujuan meyakinkan atau mengajak), dan emotif (bertujuan menyampaikan emosi atau perasaan).

Pada kelas menulis tingkat madya di lembaga pengajaran bahasa Indonesia untuk orang asing, yang menjadi tujuan akhir dari pembelajaran adalah siswa mampu membuat sebuah tulisan deskripsi. Tugas menulis yang diberikan adalah yang terkontrol (*responsive*) dan jenis tes yang bersifat *open-ended*. Pada ujian tengah semester, salah satu soal meminta siswa untuk membuat tulisan deskripsi sepanjang satu paragraf. Sementara itu, pada ujian akhir semester, siswa diminta untuk membuat sebuah esai yang terdiri dari beberapa paragraf.

2.5 Prosedur Penilaian

Untuk menilai tulisan berupa paragraf atau esai, diperlukan suatu alat penilaian yang memudahkan penilai untuk menilai tulisan itu. Menurut Brown (2004, hlm. 241), “*Three major approaches to scoring writing performance are commonly used by test designers: holistic, primary trait, and analytical.*” Ketiga prosedur penilaian yang dikemukakan oleh Brown itu dikenal juga dengan istilah rubrik, skema penilaian, atau borang.

Menurut Moskal (2000), pengertian rubrik atau skala penilaian adalah suatu metode yang digunakan untuk menilai performa murid. Mertler (2001) juga mendefinisikan rubrik sebagai panduan untuk menilai performa atau tugas murid yang terdiri dari kriteria-kriteria tertentu. Heidi Goodrich (1997) menyatakan bahwa rubrik adalah alat penilaian yang terdiri dari kriteria-kriteria yang akan diukur, seperti tujuan penulisan, organisasi, dan mekanik. Rubrik juga menunjukkan gradasi kualitas dari yang paling baik sampai yang paling buruk. Senada dengan Goodrich, Reddy (2010) juga menjelaskan bahwa “*A rubric in education literature is commonly understood as an assessment tool that is used to describe and score observable qualitative differences in performances. It captures the essence of performance in academic tasks by listing the criteria, of what counts, and describing levels of quality from excellent to poor*” (hlm. 84).

Popham (1997) dalam Reddy (2010) menyebutkan tiga unsur penting yang ada dalam sebuah skema penilaian, yaitu kriteria evaluasi (*evaluation criteria*), definisi kualitas (*quality definitions*), dan strategi penilaian (*scoring strategy*). Kriteria evaluasi adalah faktor-faktor yang dipertimbangkan oleh penilai dalam menentukan kualitas tulisan siswa. Definisi kualitas adalah penjelasan detail mengenai apa yang harus dilakukan siswa untuk menunjukkan level kemahiran yang telah dikuasai. Sementara itu, strategi penilaian sebuah skema penilaian mencakup pemakaian skema untuk menilai tulisan. Skema penilaian holistik memberikan penilaian tulisan secara menyeluruh tanpa harus menilai komponen-komponen tulisan secara terpisah. Sementara itu, skema penilaian analitik menggunakan strategi penilaian dengan memberikan nilai untuk tiap komponen tulisan yang kemudian akhirnya nilai-nilai itu menghasilkan nilai akhir.

Penggunaan skala penilaian sangat penting karena nilai yang dihasilkan akan digunakan untuk membuat keputusan atau melihat kemampuan dari siswa. Sebuah nilai yang dihasilkan dalam penilaian menulis tidak saja hasil interaksi antara peserta ujian dan tesnya, tapi juga antara peserta, tugas, penilai, dan skema penilaian yang digunakan. Dari unsur-unsur tersebut, dua hal yang berperan penting dalam penilaian adalah menjelaskan skema penilaian yang digunakan, dan memastikan si penilai menggunakannya secara benar dan konsisten (Weigle, 2002).

Menurut Moskal (2000), skema penilaian merupakan salah satu alternatif untuk menilai tulisan. Selain skema penilaian, metode *checklist* juga dapat digunakan untuk menilai tulisan. Namun, metode *checklist* lebih cocok digunakan ketika informasi yang ingin didapat sebatas menentukan apakah tulisan atau komponen tulisan sudah mencapai kriteria yang ditentukan atau belum.

2.5.1 Jenis Skala Penilaian

Sebelum menilai sebuah tulisan, yang pertama-tama harus diperhatikan adalah bagaimana cara menilainya dan jenis skala penilaian apa yang akan digunakan. Apakah sebuah tulisan diberi nilai langsung atau sebuah tulisan dinilai berdasarkan kategorinya? Coombe (2010, hlm. 183) juga menambahkan, “*An important part of writing assessment deals with selecting the appropriate writing scale.*”

Menurut Weigle (2002), ada tiga jenis skala penilaian yang dapat digunakan untuk menilai sebuah tulisan, yakni *primary trait scales*, skala holistik, dan skala analitik. Ketiga skala itu dibedakan atas dua hal: (1) apakah skala tersebut digunakan untuk tugas menulis yang spesifik atau untuk tugas yang umum (ujian kelas), dan (2) apakah sebuah tulisan akan diberikan angka tunggal (*single score*) atau multiangka (*multiple scores*).

2.5.1.1 Skala *Primary Trait*

Menurut Weigle (2002), skala *primary trait* diasosiasikan dengan apa yang dikerjakan oleh Lloyd-Jones pada 1977 untuk National Assessment of Educational Progress (NAEP), yaitu program ujian skala besar untuk sekolah-sekolah di Amerika Serikat. Ide pokok dari skala *primary trait* adalah bagaimana

siswa menulis sebaik mungkin dengan tugas yang spesifik, misalnya harus menulis jenis tulisan yang bersifat persuasif atau penjelasan. Penilaian yang diberikan dalam skala jenis ini adalah sejauh mana siswa bisa memenuhi tugas yang diberikan. Misalnya, siswa disuruh menulis tulisan persuasi, nilai yang diberikan berdasarkan sebagus apa tulisan itu dalam memberikan persuasi kepada pembaca tulisannya.

Skala *primary trait* ini sudah jelas memakan waktu dan tenaga yang cukup banyak sebab dibuat untuk satu buah penugasan. Apabila tugas yang diberikan berbeda, harus dibuat lagi skala yang baru. Lloyd-Jones (1977, dalam Weigle, 2002) memperkirakan bahwa membuat satu buah skala *primary trait* membutuhkan waktu 60 sampai 80 jam.

Dalam pembelajaran bahasa kedua, penilaian dengan skala *primary trait* tidak banyak digunakan. Namun, Hamp-Lyons (1991a) menyatakan bahwa skala jenis ini mungkin berguna bagi pemelajar bahasa kedua dalam konteks sekolah sehingga orang tua mereka—yang tidak bisa atau tidak mengerti bahasa yang digunakan di sekolah—dapat melihat perkembangan bahasa anak mereka itu.

Shaw dan Weir (2007) juga menambahkan sebuah skala *primary trait* terdiri dari (a) tugasnya, (b) pernyataan ciri-ciri yang ingin dicapai, (c) interpretasi dari hipotesis tugas performa menulis yang diharapkan, (d) penjelasan bagaimana tugas dan *primary trait* berhubungan, (e) panduan penilaian, (f) contoh sampel, dan (g) penjelasan nilai pada contoh sampel. Kriteria penilaian yang ada di dalam skala *primary trait* ini terbatas digunakan pada tugas menulis tertentu dan tidak dapat digeneralisasikan kepada tugas menulis yang lain.

2.5.1.2 Skala Holistik

Skala holistik memberikan nilai untuk sebuah tulisan berdasarkan impresi secara keseluruhan dari tulisan itu. Dalam skala holistik ini, sebuah tulisan dibaca secara cepat, dan hanya sekali, lalu diberi nilai berdasarkan skala nilai yang telah ditentukan sebelumnya (Weigle, 2002).

Skala holistik telah lama digunakan untuk penilaian tulisan dan memiliki beberapa keunggulan. Dari sudut pandang kepraktisan, dengan menggunakan skala ini, penilai jadi lebih cepat (dan juga lebih murah) dan hanya perlu membaca

tulisan satu kali lalu memberikan nilai. Keunggulan lain dari skala ini adalah skala ini membuat pembaca lebih fokus pada kekuatan dari tulisan tersebut, bukan pada kekurangannya (White, 1984 dan 1985 dalam Weigle, 2002).

White juga menambahkan bahwa skala holistik lebih andal daripada skala analitik karena ia lebih autentik; pembaca menanggapi langsung setelah membaca tulisan tersebut, sementara dalam skala analitik terlalu banyak perhatian yang diberikan sehingga mencederai makna tulisan itu secara keseluruhan.

Di sisi lain, skala jenis ini juga memiliki beberapa kekurangan, terutama dalam konteks pembelajaran bahasa kedua. Kekurangan pertama adalah nilai tunggal yang diberikan kurang memberi informasi yang jelas mengenai kemampuan menulis siswa. Juga, dengan menggunakan nilai tunggal, pengajar atau penilai tidak bisa melihat letak kekurangan siswa, apakah itu dalam hal tata bahasa, kosakata, ejaan, atau yang lainnya. Hal ini terutama bisa menjadi masalah bagi pemelajar bahasa kedua. Sebab, tiap aspek (tata bahasa, organisasi, kosakata, ejaan, dan lain-lain) dari kemampuan menulis berkembang tidak secara bersamaan. Tiap orang memiliki perkembangan yang berbeda-beda; ada yang mahir dalam kosakata dan ejaan, tetapi sangat lemah dalam tata bahasa, begitu juga sebaliknya.

Kekurangan lain dari skala holistik ini adalah nilai yang diberikan tidak selalu mudah untuk diinterpretasikan. Sebab, tiap penilai tidak selalu menggunakan kriteria yang sama dalam menilai suatu tulisan. Misalnya, sebuah tulisan diberi nilai 4 oleh seorang penilai berdasarkan komponen retorikalnya (isi, organisasi, dan pengembangan), sementara penilai yang lain juga memberi nilai 4, tetapi ia menilai berdasarkan komponen linguistiknya (penguasaan tata bahasa dan kosakata).

Coombe (2010) menyatakan bahwa skala holistik dapat dinyatakan andal dan cepat apabila satu tulisan diperiksa oleh tiga sampai empat penilai. Aturan utama dalam pemakaian skala holistik adalah seorang penilai tidak boleh menilai lebih dari dua jam dan tiap jam hanya boleh menilai maksimal 20 tulisan.

Masih menurut Coombe, beberapa keuntungan menggunakan skala holistik adalah pertama, sangat andal apabila dikerjakan tanpa batas waktu dan penilai mendapatkan pelatihan yang cukup. Kedua, skala jenis ini dirasa lebih cepat penggunaannya daripada metode penilaian tulisan yang lain, terutama

apabila jumlah tulisan yang harus diperiksa ada banyak. Ketiga, mengingat seluruh kemampuan menulis dinilai secara keseluruhan, penggunaan skala holistik tidak memperlihatkan letak kekurangan siswa, misalnya kurang bagus dalam tata bahasa. Terakhir, skala holistik lebih menunjukkan kekuatan dari tulisan daripada kelemahannya.

Namun, Coombe (2010) juga memaparkan beberapa kekurangan skala holistik. Pertama, skala holistik bisa sangat tidak andal apabila proses penilaian dilakukan dalam waktu yang terbatas dan penilai tidak memiliki pengalaman yang banyak serta belum pernah diberi pelatihan. Kedua, skala holistik tidak dapat memberikan balikan kepada siswa. Ketiga, skala holistik melihat tulisan secara keseluruhan, karena itu ada kecenderungan si penilai melupakan unsur-unsur yang membangun sebuah tulisan. Contoh dari skema holistik adalah seperti berikut.

Tabel 2.1 Skala holistik

<i>Template for Holistic Rubrics</i>	
<u>Score</u>	<u>Description</u>
5	Demonstrates complete understanding of the problem. All requirements of task are included in response.
4	Demonstrates considerable understanding of the problem. All requirements of task are included.
3	Demonstrates partial understanding of the problem. Most requirements of task are included.
2	Demonstrates little understanding of the problem. Many requirements of task are missing.
1	Demonstrates no understanding of the problem.
0	No response/task not attempted.

(Mertler, 2001, hlm. 3)

2.5.1.3 Skala Analitik

Dalam skala analitik, sebuah tulisan diberi nilai berdasarkan beberapa komponen penulisan. Tergantung dari tujuan penilaian, sebuah skala analitik bisa terdiri dari komponen isi, organisasi tulisan, kepaduan kalimat, laras bahasa, kosakata, tata bahasa, dan ejaan. Skala analitik, yang memberikan informasi detail mengenai kemampuan menulis seseorang dari berbagai sudut pandang, lebih banyak digunakan dibanding skala holistik (Weigle, 2002).

Keunggulan utama dari skala analitik dibandingkan skala holistik adalah skala ini memberikan informasi yang lebih detail mengenai kemampuan menulis siswa. Skala penilaian ini sangat bagus terhadap *washback effect*, baik terhadap murid ataupun pengajar. *Washback effect*, atau biasa disebut juga *backwash effect*, adalah efek yang ditimbulkan dari sebuah tes terhadap proses pembelajaran serta pemelajaran, dan efek tersebut bisa bersifat menguntungkan atau merugikan (Hughes, 2003).

Menurut Weigle (2002), *washback* positif adalah efek sebuah tes yang mendorong pengajar untuk menggunakan metode terbaik dalam bidang pengajaran tersebut. Sementara itu, *washback* negatif adalah efek yang ditimbulkan dari sebuah tes yang mendorong pengajar untuk menggunakan metode yang kontra-produktif, atau yang tidak sejalan dengan metode terbaik yang sedang berkembang.

Coombe (2010) mengatakan bahwa skala analitik sangat berguna digunakan oleh penilai yang kurang pengalaman. Cooper (1977) juga menambahkan bahwa skala penilaian analitik ini sangat berguna dalam evaluasi program dan untuk tujuan penelitian. Beberapa penelitian menunjukkan bahwa skala analitik lebih mudah untuk diajarkan kepada para penilai yang minim pengalaman, sebab para penilai yang belum berpengalaman lebih mudah mengerti dan menerapkan kriteria-kriteria yang diajarkan untuk penilaian. Skala analitik juga sangat membantu para pemelajar bahasa kedua karena mereka jadi tahu di mana letak kekurangan dan kelebihan mereka dalam menulis. Selain itu, nilai yang diberikan menggunakan skala analitik lebih andal dibandingkan nilai dari skala holistik.

Beberapa keuntungan dari pemakaian skala analitik antara lain, pertama skala analitik membantu pengajar dengan menampilkan kelemahan dan kelebihan siswa dalam kemahiran menulis. Kedua, seperti telah disebutkan, cukup andal walaupun penilai kurang berpengalaman dan hanya sedikit mendapat pelatihan. Terakhir, proses pelatihan penilai untuk menggunakan skala analitik lebih mudah karena skala ini sudah eksplisit dan detail. Akan tetapi, Coombe (2010) juga memaparkan satu kekurangan besar skala analitik, yaitu penggunaan skala analitik lebih memakan waktu karena penilai harus memeriksa semua unsur tulisan.

Namun begitu, skala penilaian analitik ini juga tak luput dari kekurangan. Kekurangan utama dari skala analitik adalah membutuhkan waktu lebih lama untuk memeriksa sebuah tulisan karena komponen yang akan dinilai ada lebih dari satu. Selain itu, menurut Cooper (1977), untuk membuat skala penilaian yang tepat dibutuhkan enam tahap yang harus dipertimbangkan. Pertama, fitur-fitur yang ada di dalam skala harus berasal dari sampel asli tulisan murid. Kemudian, skala tersebut diujikan terlebih dahulu. Ketiga, kualitas efektif dan tidak efektifnya juga harus dipertimbangkan. Keempat, nilai harus diberikan pada tiap kriteria. Kelima, penguji harus memakai skala yang baru dengan sampel yang baru pula. Dan terakhir, tingkat keandalan kriteria tersebut harus diperhitungkan.

Sementara itu, menurut Davies (1999), dalam penilaian tulisan, keuntungan menggunakan skala holistik dibandingkan analitik adalah proses penilaian tidak membutuhkan waktu yang lama. Namun, kelemahannya adalah mungkin fokus penilaian dari beberapa penilai itu bisa berbeda-beda. Sementara, kelebihan skala analitik adalah penilai harus fokus pada tiap kriteria yang terdapat dalam tulisan sehingga nilai dari beberapa penilai itu bisa kurang lebih sama.

Salah satu skala penilaian analitik yang baik dan banyak digunakan dalam ESL adalah seperti yang dibuat oleh Jacobs (1981). Skala yang dibuat oleh Jacobs itu memasukkan lima kriteria penilaian untuk sebuah tulisan, yaitu isi, organisasi, kosakata, penggunaan bahasa, dan mekanisasi. Skala lengkapnya adalah seperti berikut.

Tabel 2.2 Skala analitik

ESL COMPOSITION PROFILE		
STUDENT	DATE	TOPIC
SCORE	LEVEL	CRITERIA
CONTENT	27-30	EXCELLENT TO VERY GOOD: knowledgeable, substantive, thorough development of thesis, relevant to assigned topic.
	22-26	GOOD TO AVERAGE: some knowledge of subject, adequate range, limited development of thesis, mostly relevant to topic, but lacks detail.
	17-21	FAIR TO POOR: limited knowledge of subject, little substance, inadequate development of topic.
	13-16	VERY POOR: does not show knowledge of subject, non-substantive, not pertinent, OR not enough to evaluate.
ORGANIZATION	18-20	EXCELLENT TO VERY GOOD: fluent expression, ideas clearly stated/supported, succinct, well-organized, logical sequencing, cohesive.
	14-17	GOOD TO AVERAGE: somewhat choppy, loosely organized but main ideas stand out, limited support, logical but incomplete sequencing.
	10-13	FAIR TO POOR: non-fluent, ideas confused or disconnected, lacks logical sequencing and development.
	7-9	VERY POOR: does not communicate, no organization, OR not enough to evaluate.
VOCABULARY	18-20	EXCELLENT TO VERY GOOD: sophisticated range, effective word/idiom choice and usage, word form mastery, appropriate register.
	14-17	GOOD TO AVERAGE: adequate range, occasional errors of word/idiom form, choice, usage <i>but meaning not obscured</i> .
	10-13	FAIR TO POOR: limited range, frequent errors of word/idiom form, choice, usage, <i>meaning confused or obscured</i> .
	7-9	VERY POOR: essentially translation, little knowledge of English vocabulary, idioms, word form, OR not enough to evaluate.
LANGUAGE USE	22-25	EXCELLENT TO VERY GOOD: effective complex constructions, few errors of agreement, tense, number, word order/function, articles, pronouns, prepositions.
	18-21	GOOD TO AVERAGE: effective but simple constructions, minor problems in complex constructions, several errors of agreement, tense, number, word order/function, articles, pronouns, prepositions <i>but meaning seldom obscured</i> .
	11-17	FAIR TO POOR: major problems in simple/complex constructions, frequent errors of negation, agreement, tense, number, word order/function, articles, pronouns, prepositions and/or fragments, run-ons, deletions, <i>meaning confused or obscured</i> .
	5-10	VERY POOR: virtually no mastery of sentence construction rules, dominated by errors, does not communicate, OR not enough to evaluate.
MECHANICS	5	EXCELLENT TO VERY GOOD: demonstrates mastery of conventions, few errors of spelling, punctuation, capitalization, paragraphing.
	4	GOOD TO AVERAGE: occasional errors of spelling, punctuation, capitalization, paragraphing <i>but meaning not obscured</i> .
	3	FAIR TO POOR: frequent errors of spelling, punctuation, capitalization, paragraphing, poor handwriting, <i>meaning confused or obscured</i> .
	2	VERY POOR: no mastery of conventions, dominated by errors of spelling, punctuation, capitalization, paragraphing, handwriting illegible, OR not enough to evaluate.
TOTAL SCORE	READER	COMMENTS

(Jacobs, 1981)

2.6 Prinsip Praktis, Andal, dan Sahih

Sebuah tes yang baik memiliki tingkat keandalan dan kesahihan yang tinggi serta praktis untuk diterapkan. Menurut Brown (2004), sebuah tes yang praktis adalah yang tidak terlalu mahal, tidak memakan waktu yang lama untuk dikerjakan atau sesuai dengan waktu yang diberikan, mudah untuk dilaksanakan, serta mempunyai prosedur penilaian yang spesifik dan efisien dalam waktu. Sementara itu, sebuah tes dikatakan tidak praktis antara lain apabila memakan waktu yang lama dalam pengerjaannya atau pemeriksaannya. Sebuah tes yang dapat diselesaikan oleh siswa hanya dalam beberapa menit dan membutuhkan waktu berjam-jam untuk dinilai oleh pengajar juga dikatakan tidak praktis.

Menurut Brown (2004), sebuah tes yang andal adalah yang konsisten dan memberikan hasil yang sama pada ujian atau percobaan yang berulang. Jika Anda memberikan sebuah tes yang sama kepada seorang murid dalam dua kesempatan yang berbeda, dan tidak ada pelatihan di antara kedua tes tersebut, hasil dari tes itu harus menghasilkan skor yang sama jika memang tes itu andal. Ada beberapa faktor yang memengaruhi keandalan sebuah tes, yakni peserta tes (*student-related reliability*), penilai tes (*rater reliability*), penyelenggaraan tes (*test administration reliability*), dan tes itu sendiri (*test reliability*).

Ada beberapa hal yang memengaruhi peserta tes, antara lain kesehatan, kelelahan, kecemasan, serta faktor fisik dan faktor psikologis lain. Hal-hal tersebut dapat memengaruhi siswa dalam menyelesaikan tes. Sementara itu, subjektivitas, *human error*, dan kelelahan bisa menjadi faktor yang memengaruhi penilai.

Keandalan erat kaitannya dengan konsistensi (Nunan dan Bailey, 2009). Menurut Brown (2004), ada dua macam keandalan penilai, yakni keandalan antarpemilai (*inter-rater reliability*) dan keandalan intrapemilai (*intra-rater reliability*). Apabila ada dua orang penilai memeriksa sebuah tulisan yang sama dan memberikan nilai yang sama, mereka dikatakan memiliki keandalan antarpemilai. Ada beberapa hal yang bisa memengaruhi keandalan antarpemilai, di antaranya kurang mengerti akan kriteria penilaian, kurang berpengalaman, atau kurang perhatian. Sementara itu, keandalan intrapemilai adalah mengenai keandalan pada diri si penilai itu sendiri. Seorang penilai dikatakan memiliki

keandalan intrapenilai apabila ia memeriksa sebuah tulisan sebanyak dua kali—dengan jeda waktu, dan nilai yang dihasilkan dari kedua penilaian itu sama. Keandalan intrapenilai ini erat kaitannya dengan konsistensi penilai dalam menerjemahkan prosedur penilaian.

Nunan dan Bailey (2009) juga menambahkan bahwa ada dua jenis keandalan, yaitu keandalan internal dan keandalan eksternal. Keandalan biasanya berhubungan dengan replikasi atau pengulangan. Jika kita mengumpulkan data sebanyak dua kali (dari murid yang sama, yang tidak memperoleh pelajaran apa pun selama rentang waktu itu) dan keduanya mendapatkan hasil yang sama, kita bisa mengklaim bahwa data Anda andal, dan penelitian Anda memiliki keandalan internal. Sementara itu, jika ada orang lain yang melakukan penelitian dengan subjek yang sama dan menggunakan standar yang telah teruji, serta mendapatkan hasil yang sama juga, Anda bisa menyatakan bahwa penelitian Anda memiliki keandalan eksternal.

Ketidakandalan juga bisa muncul karena faktor penyelenggaraan tes. Ketika tes berlangsung, suara bising dari luar kelas, kondisi meja dan kursi, kondisi foto kopi soal, serta faktor kurangnya cahaya juga dapat memengaruhi keandalan. Apabila sebuah tes berlangsung dalam waktu yang sangat lama, siswa bisa saja kelelahan dan hilang konsentrasi ketika mengerjakan soal-soal yang ada di bagian akhir. Waktu ujian yang terlalu singkat juga dapat memberikan tekanan kepada siswa sehingga mereka mengerjakan soal dengan tidak maksimal dan terburu-buru.

Sebuah tes dengan kesahihan yang baik adalah yang sesuai dengan materi atau kemampuan yang akan diukur. Menurut Nunan dan Bailey (2009), kesahihan ada dua, yaitu kesahihan internal dan kesahihan eksternal. Kesahihan internal berhubungan dengan apakah nilai dari sebuah tes memang benar hasil dari *treatment* atau pelajaran yang diberikan. Sementara itu, sebuah studi dikatakan memiliki kesahihan eksternal apabila hasil penemuannya yang berdasarkan sampel itu bisa digeneralisasikan kepada populasi yang lebih luas.

Menurut Brown (2004), kesahihan ada empat jenis, yakni kesahihan isi (*content validity*), kesahihan kriteria (*criterion-related validity*), kesahihan konsekuensi (*consequential validity*), kesahihan wajah (*face validity*), dan

kesahihan konstruk (*construct validity*). Sebuah tes dikatakan memiliki kesahihan isi apabila tes yang diberikan tersebut memang sesuai untuk mengukur atau menguji materi-materi yang telah dipelajari oleh siswa. Menurut Brown (2004), hal terpenting untuk mencapai kesahihan isi dalam penilaian di sebuah kelas adalah dengan menguji performa murid secara langsung. Karena itu, jika pengajar ingin menguji kemampuan menulis seorang murid, tentu ia harus meminta si murid membuat sebuah tulisan.

Kesahihan kriteria dapat ditunjukkan dengan cara membandingkan hasil dari sebuah penilaian dengan hasil dari ujian lain dengan kriteria/materi yang mirip. Kesahihan kriteria memiliki dua kategori, yakni kesahihan bersama (*concurrent validity*) dan kesahihan pemrakira (*predictive validity*). Kesahihan pemrakira sangat penting dalam tes penempatan, tes bakat bahasa (*language aptitude test*), dan sejenisnya. Kesahihan pemrakira bukan untuk mengukur kemampuan secara bersamaan, tapi untuk mengukur dan memprediksikan kesuksesan siswa di masa yang akan datang. Sementara itu, sebuah tes dikatakan memiliki kesahihan bersama jika nilai dari tes itu didukung dengan pembuktian performa secara bersamaan. Contohnya adalah, jika seorang siswa mendapat nilai tinggi dalam ujian akhir pembelajaran bahasa asing, tentu harus dibuktikan dengan bentuk nyata ketika berkomunikasi.

Kesahihan konsekuensi meliputi hal-hal yang terkait dengan tes itu, misalnya ketepatan dalam mengukur kemampuan yang diinginkan, pengaruhnya dalam persiapan ujian, pengaruhnya terhadap pemelajar, dan lainnya. Dampak lain yang lebih penting dari kesahihan konsekuensi adalah yang disebut *washback effect*.

Kesahihan wajah lebih berkaitan dengan pandangan siswa akan tes tersebut, apakah tes itu sudah relevan, adil, dan berguna dalam meningkatkan proses pembelajaran (Gronlund, 1998, dalam Brown, 2004). Kesahihan wajah dikatakan sudah tepat apabila siswa melihat bahwa (1) tes itu sudah dibuat dengan benar dan tugasnya sesuai dengan apa yang pernah diajarkan, (2) tes itu dapat dikerjakan dalam jangka waktu yang telah diberikan, (3) soal-soalnya cukup jelas dan tidak membingungkan, (4) petunjuk yang jelas, dan (6) tingkat kesulitan yang sesuai dengan tingkat kemahiran mereka.

Kroll (1998) menyatakan bahwa dua hal penting yang harus ada dalam pengtesan instrumen penilaian tulisan adalah kesahihan dan keandalan. Ada beberapa jenis kesahihan, tetapi yang paling penting dalam tes menulis adalah kesahihan konstruk. Menurut Kroll (1998), kesahihan konstruk harus mengetes apa yang dibutuhkan untuk mencapai kesuksesan dalam kemahiran menulis. Sementara itu, menurut Brown (2004, hlm. 25), *“A construct is any theory, hypothesis, or model that attempts to explain observed phenomena in our universe of perceptions.”*

Menurut Weigle (2002, hlm. 121), secara kesahihan konstruk, skala analitik *“more appropriate for L2 writers as different aspects of writing ability develop at different rates.”* Sementara itu, skala holistik *“assume that all relevant aspects of writing ability develop at the same rate and can thus be captured in a single score; correlate with superficial aspects such as length and handwriting.”* Karena itu, berdasarkan apa yang telah dikemukakan oleh Weigle, pemilihan skala analitik untuk menilai tulisan dirasa sudah tepat secara kesahihan konstruk.

2.7 Landasan Teori

Jenis tulisan yang akan menjadi sampel pengujian skema penilaian berupa tulisan esai dan paragraf. Tulisan yang menjadi sampel penelitian ini dihasilkan siswa melalui sebuah tes langsung, yakni sebuah cara penilaian kemampuan menulis siswa dengan melihatnya dari tulisan yang dihasilkan. Penulis mengacu ke teori yang dikemukakan oleh Coombe (2010) yang menyatakan bahwa dalam membuat tulisan itu siswa harus memerhatikan mengenai isi tulisan, organisasi ide, serta menggunakan kosakata, tata bahasa, dan sintaksis yang tepat. Teori Coombe tersebut juga menjadi acuan bagi penulis dalam menentukan kriteria-kriteria penilaian yang akan dimasukkan ke skema penilaian yang akan dirancang.

Sementara itu, tulisan yang dihasilkan siswa merupakan bentuk dari penilaian sumatif. Menurut Brown (2004), penilaian sumatif berfungsi mengukur pemahaman siswa atas suatu pelajaran atau unit instruksi. Oleh karena itu, teori dari Brown itu juga menjadi acuan bagi penulis untuk menghasilkan sebuah skema penilaian yang cocok untuk penilaian sumatif. Dari skema penilaian yang

penulis rancang, diharapkan pengajar dan siswa mendapatkan balikan yang dapat dijadikan pegangan sebagai perbaikan ke depannya.

Untuk menilai tulisan, baik itu berupa esai atau hanya sebuah paragraf, dibutuhkan sebuah skema penilaian untuk memudahkan proses penilaian. Penulis mengacu kepada teori yang dikemukakan oleh Weigle (2002) mengenai jenis-jenis skema penilaian. Dari uraian Weigle (2002) mengenai jenis-jenis skema penilaian, peneliti memilih untuk merancang skema penilaian analitik karena beberapa keunggulan yang dimiliki skala analitik dibandingkan skala penilaian yang lain dalam menilai sebuah tulisan dan memberikan balikan.

Dari teori-teori yang telah diuraikan di atas, peneliti memakai teori dari Brown (2004) mengenai prinsip keandalan dan kepraktisan. Prinsip keandalan dari Brown yang penulis jadikan sebagai landasan teori adalah mengenai keandalan antarpemilai. Jika nilai yang diberikan oleh ketiga partisipan terhadap tulisan yang sama memiliki perbedaan yang tidak signifikan, dapat dikatakan bahwa skema penilaian yang penulis adaptasi dari Jacobs (1981) ini sudah memenuhi prinsip keandalan antarpemilai.

Prinsip kepraktisan terkait dengan proses penggunaan skema penilaian yang akan digunakan dalam menilai tulisan. Skema penilaian yang dihasilkan dalam penelitian ini dapat dikatakan praktis digunakan apabila memenuhi syarat-syarat kepraktisan menurut Brown (2004), yaitu tidak terlalu mahal, tidak memakan waktu lama untuk dikerjakan, mudah untuk dilaksanakan, dan mempunyai prosedur penilaian yang spesifik.

Sementara itu, penulis menggunakan teori dari Weigle (2002) mengenai prinsip kesahihan konstruk sebagai acuan dalam memilih jenis skema penilaian yang akan dibuat. Prinsip kesahihan konstruk menjadi salah satu dasar bagi penulis untuk memilih skema penilaian berjenis analitik dibandingkan skala holistik.

BAB 3

METODE PENELITIAN

Nunan (1992) menjelaskan bahwa penelitian adalah suatu proses penyelidikan sistematis yang terdiri dari tiga komponen, yaitu (1) pertanyaan, masalah, atau hipotesis, (2) data, dan (3) analisis dan interpretasi. Nunan dan Bailey (2009) menyebutkan bahwa penelitian berbasis kelas terbagi menjadi dua jenis, yakni *psychometric* dan *naturalistic*. Penelitian *psychometric* disebut juga penelitian kuantitatif, sementara penelitian *naturalistic* disebut juga penelitian kualitatif.

Penelitian kuantitatif dan kualitatif masing-masing mewakili cara pengumpulan dan analisis data yang berbeda-beda. Akan tetapi, Allwright dan Bailey (1991) menyatakan bahwa data penelitian kelas dapat dikumpulkan dan dianalisis secara kualitatif maupun kuantitatif. Dalam penelitian kuantitatif, pada umumnya peneliti menyelidiki apakah metode, materi, atau cara pengajaran yang berbeda juga menghasilkan perbedaan dalam pembelajaran bahasa. Sementara itu, tujuan dari penelitian kualitatif adalah untuk mendapatkan wawasan mengenai kompleksitas pembelajaran dan pemelajaran bahasa melalui proses observasi dan deskripsi (Nunan dan Bailey, 2009). Grotjahn (1987, dalam Nunan dan Bailey, 2009) menyatakan bahwa perbedaan dalam penelitian kuantitatif dan kualitatif dibedakan dalam tiga hal, yaitu desain (apakah penelitian berbasis eksperimen, *quasi-experimental*, atau non-eksperimen), jenis data yang dikumpulkan, dan jenis analisis yang dilakukan. Penelitian yang penulis lakukan ini merupakan penelitian kualitatif dan kuantitatif, sebab penulis melakukan penelaahan data berupa hasil wawancara dan data angka-angka yang didapat dari hasil pengujian skema penilaian oleh para partisipan.

3.1 Populasi dan Sampel

Butler (1985) menyatakan bahwa populasi biasanya merujuk pada sekumpulan wujud, sekumpulan orang, atau sekumpulan makhluk hidup. Namun, ia menyatakan bahwa dalam statistika istilah itu digunakan secara lebih umum, yaitu dapat merujuk ke bentuk dan jenis apa pun yang menjadi sasaran penyelidikan.

Populasi penelitian ini adalah karya tulis dari siswa kelas menulis bahasa Indonesia untuk orang asing tingkat madya periode Januari-Mei 2011. Penelitian ini hanya akan merancang skema penilaian untuk tingkat madya karena pada tingkat ini siswa sudah dapat menghasilkan sebuah tulisan panjang berupa esai sehingga skema penilaian yang digunakan pun bisa diterapkan untuk semua jenis tulisan pada tingkat ini. Sementara itu, di tingkat akhir, kelas menulis sudah terbagi menjadi dua, yaitu kelas menulis karya populer dan kelas menulis karya ilmiah. Oleh karena itu, skema penilaian yang digunakan untuk kedua kelas itu pun berbeda. Sementara itu, alasan penulis tidak meneliti di tingkat dasar adalah karena pada tingkat ini siswa belum mampu atau belum diajarkan untuk membuat tulisan yang panjang mengingat kosakata mereka masih terbatas. Produksi tulisan pada tingkat dasar hanya sebatas membuat kalimat sederhana sehingga proses menilai tulisannya pun tidak memerlukan skema penilaian.

Kelas menulis madya pada periode Januari-Mei 2011 ada enam kelas. Sementara itu, karya tulis yang menjadi populasi penelitian adalah hasil Ujian Tengah Semester (UTS) dan hasil Ujian Akhir Semester (UAS). Sampel yang peneliti ambil berjumlah 36 buah. Perinciannya adalah dari setiap kelas itu diambil satu orang murid dengan kemampuan yang pandai, satu orang murid dengan kemampuan sedang, dan satu orang murid dengan kemampuan yang kurang. Jadi, dari enam kelas menulis itu, ada 18 murid yang dijadikan sampel penelitian. Penentuan mana murid dengan kemampuan pintar, sedang, atau kurang berdasarkan nilai harian dan pengamatan dari masing-masing pengajar di kelas tersebut. Kemudian, hasil UTS dan UAS ke-18 murid itulah yang akan diuji menggunakan skema penilaian yang peneliti akan buat sehingga akan mendapatkan data yang siap untuk dianalisis.

3.2 Partisipan Penelitian

Peneliti meminta bantuan partisipan untuk menilai tulisan menggunakan skema penilaian yang telah dibuat. Karena tulisan yang akan dinilai diambil dari kelas menulis tingkat madya, partisipan yang dipilih oleh peneliti adalah yang memiliki pengalaman mengajar di kelas menulis tingkat madya. Jumlah partisipan

ini ada tiga orang, dan ketiganya adalah perempuan. Selanjutnya ketiga partisipan ini akan disebut dengan Partisipan 1, Partisipan 2, dan Partisipan 3.

Partisipan 1 berusia 26 tahun dan memiliki pengalaman mengajar bahasa Indonesia untuk orang asing kurang lebih tiga tahun. Di antara ketiga partisipan, Partisipan 1 merupakan yang termuda dari segi usia dan pengalaman mengajar bahasa Indonesia untuk orang asing. Partisipan 1 memiliki pengalaman mengajar kelas menulis bahasa Indonesia untuk orang asing tingkat madya selama dua tahun.

Partisipan 2 berusia 28 tahun dan memiliki pengalaman mengajar bahasa Indonesia untuk orang asing selama kurang lebih enam tahun. Walaupun tidak tiap semester mengajar kelas menulis tingkat madya, tetapi dalam setahun minimal Partisipan 2 mengajar kelas menulis tingkat madya satu kali. Bahkan, ada kalanya dalam satu tahun Partisipan 2 mengajar kelas menulis tingkat madya tiga semester berturut-turut. Sebagai informasi, di tempat Partisipan 2 bekerja dalam satu tahun ada tiga semester masa perkuliahan, yaitu periode Januari-Mei, Mei-Juli, dan Agustus-Desember. Partisipan 2 merupakan yang paling senior di antara ketiga partisipan.

Partisipan yang terakhir, yaitu Partisipan 3, berusia 27 tahun dan memiliki pengalaman mengajar bahasa Indonesia untuk orang asing kurang lebih lima tahun. Pengalaman mengajar kelas menulis bahasa Indonesia untuk orang asing tingkat madya Partisipan 3 juga sekitar lima tahun. Partisipan 3 juga tidak setiap semester mengajar kelas menulis tingkat madya, tetapi dalam satu tahun minimal mengajar kelas menulis tingkat madya satu kali.

3.3 Teknik Pengumpulan Data

Data dalam penelitian ini yang akan dianalisis merupakan angka-angka dari hasil penilaian sampel. Oleh karena itu, untuk mengumpulkan data berupa angka-angka tersebut, penulis meminta bantuan partisipan untuk mengujikan skema penilaian menggunakan sampel UTS dan UAS siswa kelas menulis madya.

Dalam proses pengumpulan data itu, masing-masing partisipan menilai 36 tulisan menggunakan skema penilaian yang penulis rancang. Nilai dari setiap tulisan itu dituliskan pada kolom yang ada di bagian bawah skema penilaian.

Nilai-nilai tersebutlah yang kemudian penulis tabulasi untuk dihitung secara statistik.

Setelah dihitung secara statistik, muncul data berupa angka-angka. Angka-angka tersebutlah yang akan dijadikan sumber pembahasan dalam penelitian ini. Selain itu, dari angka-angka itu pula peneliti dapat melakukan tindak lanjut terhadap penelitian ini.

Sementara itu, data kualitatif diperoleh dari hasil wawancara ketiga partisipan mengenai kendala dan kelebihan skema penilaian yang penulis rancang. Wawancara kepada ketiga partisipan dilakukan secara terpisah, dan di antara ketiga partisipan tidak memberi tahu yang lain mengenai isi wawancara yang dilakukan. Penulis juga melakukan wawancara informal dengan beberapa pengajar kelas menulis bahasa Indonesia untuk orang asing tingkat madya. Wawancara tersebut dilakukan sebelum penelitian ini berlangsung dengan tujuan mendapatkan informasi mengenai sebuah skema penilaian yang mudah dipakai.

3.4 Data

Data yang terkumpul dari hasil uji skema penilaian menggunakan sampel adalah data kuantitatif berupa angka-angka sehingga akan dianalisis secara statistik. Melalui perhitungan statistik, peneliti menentukan apakah skema penilaian yang digunakan sudah memiliki keandalan yang tinggi atau belum.

Selain itu, untuk mendukung atau sebagai tindak lanjut terhadap hasil analisis data kuantitatif, peneliti juga mengumpulkan data kualitatif. Data kualitatif itu berupa hasil wawancara kepada ketiga partisipan.

3.5 Penyusunan Skema Penilaian

Dalam penelitian ini, perancangan skema penilaian untuk kelas menulis pada tingkat menengah dibuat dengan mempertimbangkan syarat penyusunan sebuah penilaian yang baik, yakni keandalan, kesahihan, dan kepraktisan. Berdasarkan pertimbangan itu, penulis membuat sebuah skema penilaian dengan terlebih dahulu mengadaptasi skema penilaian dari Jacobs (1981) yang kemudian diterapkan menggunakan langkah-langkah pembuatan skema penilaian yang dikemukakan oleh Mertler (2001).

Menurut Mertler, hal pertama yang harus dipikirkan adalah tujuan dari pembuatan skema penilaian itu. Selanjutnya, pengajar harus menentukan apakah performa atau tulisan yang dibuat murid akan dinilai secara holistik atau analitik. Yang dimaksud dengan penilaian secara holistik adalah sebuah tulisan dinilai berdasarkan impresi dari penilai terhadap tulisan itu, dan nilai yang diberikan pun berupa nilai tunggal. Sementara itu, penilaian secara analitik dilakukan dengan menilai semua unsur-unsur yang membentuk sebuah tulisan. Setiap unsur itu diberi nilai tersendiri, lalu nilai-nilai itu dijumlahkan untuk mendapatkan nilai akhir.

Terlepas dari jenis skema penilaian yang dipilih, kriteria performa dan indikator lain harus diidentifikasi sebagai langkah pertama perancangan skema penilaian. Setelah mengetahui akan menilai secara holistik atau analitik, hal yang selanjutnya adalah menentukan kriteria-kriteria apa saja yang akan dinilai dan ada berapa jumlah kriteria itu. Setelah mengetahui jumlah kriteria dan jenis kriteria apa yang akan dimasukkan ke dalam skala penilaian, setiap kriteria tersebut dideskripsikan untuk memudahkan para pengajar dalam membedakan kriteria yang satu dengan kriteria yang lain serta diberikan bobot penilaiannya.

Setelah skema penilaian tahap awal terbentuk, harus diujikan terlebih dahulu untuk memeriksa apakah skema penilaian itu sudah praktis dan tidak menimbulkan kebingungan. Proses pengujian dilakukan kepada dua orang pengajar kelas menulis tingkat madya untuk mengetahui kekurangan dari skema penilaian ini, terutama dalam hal deskripsi dari setiap level. Proses pengujian itu tidak dilakukan secara bersamaan. Kedua penilai mendapatkan penjelasan terlebih dahulu dari peneliti mengenai tampilan dan isi skema penilaian. Setelah diujikan, tahap selanjutnya adalah merevisi skema penilaian berdasarkan masukan-masukan yang ada.

Berikut ini adalah langkah-langkah pembuatan skema penilaian selengkapnya yang telah dikumpulkan dari berbagai sumber oleh Mertler (2001).

Langkah 1: Tentukan tujuan pembelajaran yang akan dinilai.

Langkah 2: Identifikasi hal apa saja yang ingin diamati (begitu juga yang tidak ingin diamati) dari produk, proses, atau performa siswa.

Langkah 3: Jelaskan karakteristik yang menjelaskan setiap kriteria. Tentukan cara-cara untuk menjelaskan apa yang dimaksud dengan nilai yang di atas rata-rata, rata-rata, atau di bawah rata-rata.

Langkah 4: Tulis deskripsi untuk masing-masing kriteria penilaian dari yang tertinggi, sedang, dan terendah.

Langkah 5: Kumpulkan sampel dari tugas siswa.

Langkah 6: Revisi skala jika memang diperlukan.

Langkah-langkah pembuatan skema penilaian yang telah diuraikan oleh Mertler (2001) di atas kemudian diterapkan oleh penulis dalam penelitian ini. Namun, penulis menambahkan satu langkah terakhir dalam pembuatan skema penilaian ini, yaitu proses pengujian skema penilaian. Penjelasan langkah-langkah pembuatan skema penilaian milik penulis diuraikan lebih lanjut pada subbab-subbab berikut.

3.5.1 Penentuan Tujuan Pembuatan Skema Penilaian

Secara umum, pembuatan skema penilaian bertujuan mempermudah penilai untuk menilai tulisan siswa yang berupa esai. Menurut Goodrich (1997), skema penilaian berguna bagi pembelajar dan pemelajar dalam banyak hal. Pertama, skema penilaian dapat meningkatkan performa pemelajar dengan memperjelas ekspektasi apa yang diharapkan oleh pengajar dan menunjukkan bagaimana cara memenuhi ekspektasi itu. Kedua, skema penilaian membantu pemelajar dalam menilai kualitas tulisan mereka dan tulisan orang lain. Ketika skema penilaian digunakan untuk *self-assessment* dan *peer-assessment*, pemelajar jadi mengetahui letak kekurangan mereka dan berusaha untuk memperbaikinya. Ketiga, skema penilaian menghemat waktu pengajar dalam memeriksa tulisan siswa. Keempat, skema penilaian memudahkan pengajar karena dapat digunakan untuk kelas yang heterogen. Reddy (2010) menambahkan bahwa dalam penilaian sebuah tulisan seperti karya proyek, analisis kasus, esai, dan portofolio, yang tidak dapat dinilai dengan sepenuhnya objektif, pemakaian skema penilaian sangat membantu dalam mencapai keandalan (konsistensi) dan kesahihan (akurasi) penilaian performa siswa.

Senada dengan alasan-alasan yang dikemukakan oleh Goodrich (1997) dan Reddy (2010), pembuatan skema penilaian dalam penelitian ini adalah untuk mempermudah pengajar dalam penilaian tulisan. Selain itu, penggunaan skema penilaian juga membantu siswa mendapatkan balikan untuk mengetahui kelebihan dan kekurangan mereka. Dari skema penilaian yang telah dikembalikan kepada siswa itu, diharapkan mereka dapat memperbaiki tulisan mereka.

3.5.2 Penentuan Skema Penilaian dan Kriteria Penilaian

Setelah menentukan tujuan dari pembuatan skema penilaian, tahap yang kedua adalah menentukan jenis skema penilaian yang akan digunakan. Ada dua jenis skema penilaian yang umum digunakan untuk penilaian tulisan, yaitu skala holistik dan analitik. Skala holistik memberikan gambaran menyeluruh tentang suatu tulisan, sedangkan skala analitik menjelaskan secara terperinci hal apa saja yang membentuk sebuah tulisan. Kedua jenis skema penilaian tersebut memiliki kelebihan dan kekurangan masing-masing. Skala holistik sangat menguntungkan bagi penilaian yang memiliki waktu terbatas, sedangkan skala analitik sangat menguntungkan dalam pemberian balikan kepada pengajar maupun siswa.

Pada penelitian ini, penulis menentukan jenis skema penilaian yang akan dibuat adalah skala analitik. Alasan pemilihan skala analitik salah satunya terkait dengan tujuan dari pembuatan skema penilaian ini, yaitu membantu siswa mendapat balikan untuk mengetahui letak kelebihan dan kekurangan mereka. Dalam hal ini, skala analitik lebih baik dalam memberikan balikan karena setiap unsur-unsur yang ada dalam tulisan dinilai secara terpisah. Selain itu, seperti yang dikemukakan oleh Weigle (2002), secara kesahihan konsep skala analitik lebih cocok digunakan untuk pemelajar bahasa kedua karena tiap aspek kemahiran menulis berkembang tidak secara bersamaan. Sementara itu, skala holistik menganggap semua aspek kemahiran menulis berkembang secara bersamaan karena itu dapat dinilai secara menyeluruh.

Weir (2005) menyebutkan bahwa kriteria relevansi, komposisi atau organisasi, kepaduan, keakuratan tata bahasa, ejaan, dan tanda baca merupakan yang paling cocok untuk penilaian tugas menulis. Dari kriteria-kriteria itu, dua yang pertama dianggap yang sangat penting, sedangkan dua kriteria terakhir

dianggap kurang penting. Sementara itu, kriteria-kriteria sisanya dianggap memiliki tingkat kepentingan yang medium (Weir, 2005).

Common European Framework (*Council of Europe*, 2001) menyarankan empat sampai lima kriteria sudah cukup untuk digunakan dalam satu buah skema penilaian. Jika menggunakan sampai tujuh kriteria, itu sudah melebihi batas dan dapat menyebabkan penilai justru tidak maksimal dalam memberikan penilaian. Namun, pemakaian lima sampai enam kriteria masih dapat ditolerir oleh Common European Framework.

Pada pembuatan skema penilaian pada penelitian ini, penulis mengadaptasi bentuk skema penilaian analitik dari Jacobs (1981), terutama dalam hal kriteria penilaian dan pembagian skala. Akan tetapi, penulis tidak 100 persen meniru skema penilaian itu. Jumlah kriteria penilaian yang ada dalam skema penilaian dari Jacobs ada lima, sementara penulis menambahkan satu kriteria lagi, yakni kriteria “Capaian Tugas”. Alasannya adalah penilai juga perlu untuk menentukan apakah tulisan yang dibuat oleh siswa sudah sesuai dengan instruksi atau topik yang diberikan pada tes. Keenam kriteria yang ada dalam skema penilaian yang akan penulis buat adalah “Penggunaan Bahasa”, “Organisasi”, “Gagasan dan Kreativitas”, “Capaian Tugas”, “Kosakata”, dan “Ejaan dan Tanda Baca”.

Pada kriteria penilaian “Penggunaan Bahasa”, penekanan penilaian diberikan terhadap penggunaan kalimat sederhana dan kompleks. Selain itu, pembentukan kata menggunakan afiksasi serta penggunaan konjungsi, kata ganti, dan kata depan juga menjadi bahan penilaian. Kriteria “Penggunaan Bahasa” ini dibagi menjadi empat skala dengan menggunakan kata-kata *sedikit kesalahan*, *banyak kesalahan*, dan *tidak menguasai* untuk membedakan deskripsi yang diberikan pada tiap skala. Hal tersebut juga untuk memudahkan para penilai dalam menentukan skala nilai dari tulisan.

Kriteria penilaian yang kedua, yaitu “Gagasan dan Kreativitas” memberikan penekanan nilai terhadap pengembangan gagasan serta penggunaan contoh-contoh, ilustrasi, data, dan fakta dalam sebuah tulisan. Kriteria penilaian ini juga terbagi dalam empat skala. Untuk membedakan gradasi pencapaian dari

setiap skala, digunakan kata-kata *melebihi standar*, *sesuai standar*, dan *tidak sesuai standar*.

Kriteria penilaian “Organisasi” menekankan penilaian terhadap organisasi tulisan serta cara penyampaian ide, apakah disampaikan dengan ringkas dan jelas atau disampaikan dengan cara yang rumit dan kurang jelas. Kriteria penilaian ini juga terbagi menjadi empat skala. Kriteria penilaian “Kosakata” memberi penekanan nilai terhadap keragaman kosakata yang digunakan dalam tulisan, pemakaian kata yang tepat, dan penggunaan ragam tulisan yang tepat. Kriteria penilaian “Kosakata” juga dibagi menjadi empat skala. Kriteria penilaian “Capaian Tugas” menekankan penilaian pada relevansi tulisan dengan topik yang diberikan, relevansi tugas dengan instruksi, serta wawasan pengetahuan siswa mengenai topik yang diberikan. Kriteria penilaian yang terakhir adalah “Ejaan dan Tanda Baca.” Kriteria penilaian ini memberi penekanan nilai terhadap penguasaan ejaan, tanda baca, dan penggunaan huruf kapital yang baik. Kriteria penilaian “Ejaan dan Tanda Baca” ini dibagi menjadi tiga skala.

3.5.3 Penentuan Skala Penilaian dan Deskripsi Naratif

Penentuan jumlah skala penilaian terkait dengan bentuk skema penilaian dan kriteria yang digunakan. Andrade (1997) dan Underhill (1987) sependapat bahwa skema penilaian yang baik tidak perlu memasukkan terlalu banyak tingkatan skala. Mereka juga sependapat bahwa skala penilaian tersebut sebaiknya berjumlah genap. Alasan jumlah yang genap ini karena dapat menghilangkan kemungkinan dari penilai untuk tidak berpihak pada suatu deskripsi tertentu. Penggunaan jumlah yang genap juga untuk menghindari penilai yang ingin memberikan nilai “jalan tengah”.

Pada pembuatan skema penilaian ini, penulis menentukan semua kriteria penilaian terbagi atas empat skala, kecuali dalam kriteria “Ejaan dan Tanda Baca” yang hanya memiliki tiga skala. Alasan penentuan tiga skala untuk kriteria “Ejaan dan Tanda Baca” adalah karena rentang nilai pada kriteria itu hanya sedikit, yakni satu sampai sepuluh. Untuk membagi sepuluh angka menjadi empat skala, tentu angka pada tiap skala akan sangat kecil. Karena itu, penulis beranggapan akan

lebih baik kalau kriteria ini hanya dibagi menjadi tiga skala dengan rentang nilai tiap kriteria antara tiga sampai empat angka.

Untuk pembagian bobot nilai tiap kriteria, penulis membaginya berdasarkan tingkat kepentingan dari masing-masing kriteria dilihat dari tujuan pembelajaran di kelas menulis tingkat madya. Penulis juga melakukan wawancara dengan ketua kelas menulis di lembaga pengajaran bahasa Indonesia untuk orang asing mengenai pembagian bobot penilaian. Kriteria penilaian dengan bobot tertinggi adalah “Penggunaan Bahasa”, yakni 25 poin. Selanjutnya, berturut-turut adalah “Organisasi” (20 poin), “Gagasan dan Kreativitas” (15 poin), “Kosakata” (15 poin), “Capaian Tugas” (15 poin), dan “Ejaan dan Tanda Baca” (10 poin).

Pada tingkat madya, penggunaan tata bahasa yang baik serta penyampaian pesan dalam tulisan merupakan unsur terpenting. Proses penyampaian pesan dalam sebuah tulisan didukung dengan organisasi tulisan yang baik serta pengembangan gagasan utama dan gagasan pendukung yang tepat. Hal itu sesuai dengan pendapat dari Brown (2004) mengenai jenis tes tulisan *responsive* dan *extensive* seperti yang telah diuraikan pada subbab 2.4.3. Berdasarkan hal tersebut, kriteria “Penggunaan Bahasa”, “Organisasi”, serta “Gagasan dan Kreativitas” mendapatkan bobot yang cukup besar. Sementara itu, kriteria penilaian “Ejaan dan Tanda Baca” mendapat bobot yang kecil karena pada tingkat madya siswa diharapkan sudah menguasai ejaan dan tanda baca dengan cukup baik, sebab kemahiran ejaan dan tanda baca sudah diajarkan pada tingkat dasar dengan bobot yang cukup besar.

Sementara itu, untuk pembagian rentang nilai di setiap skala, penulis membaginya seperti bentuk piramida, yakni poin lebih banyak terdistribusi di skala bawah, kemudian semakin ke atas poin semakin sedikit terdistribusi. Menurut Luoma (2004), semakin banyak skala dalam sebuah skema penilaian, balikan yang dihasilkan akan semakin spesifik dan lebih menunjukkan perkembangan kebahasaan siswa.

Skema penilaian berkaitan dengan pengukuran, karena itu sangat penting untuk menanyakan kepada penilai berapa banyak level yang dapat dibedakan secara konsisten. Cara termudah untuk mengeceknya misalnya dengan melihat seberapa baik si penilai jika harus menilai performa yang sama dengan jeda waktu

satu minggu. Selain itu, cara yang lain adalah dengan melihat seberapa baik dua orang penilai memiliki pandangan yang sama (Luoma, 2004). Penulis melakukan pengecekan level ini dengan cara yang kedua, yakni menggunakan dua orang penilai untuk memeriksa satu tulisan yang sama.

Menurut Taufik (2009), penulisan deskripsi sebaiknya menggunakan kata-kata yang tidak mengandung lebih dari satu makna. Sering terjadi, penulisan deskripsi untuk skema penilaian menjadi sangat tidak jelas karena pembuat skema penilaian menjadi terlalu kreatif. Penggunaan kata positif dalam deskripsi juga sangat dianjurkan.

Hal lain yang perlu dipertimbangkan adalah penempatan posisi dari skala tingkatan. Penempatan posisi yang umum adalah dengan menempatkan skala tingkatan tertinggi atau terbaik di sebelah kiri atau atas skema penilaian untuk kemudian dilanjutkan dengan deskripsi berikutnya di sebelah kanan atau di bawah. Penempatan tersebut dilakukan karena kebiasaan membaca secara umum dimulai dari tingkat yang tertinggi ke yang terendah atau dari kiri ke kanan (Taufik, 2009).

3.5.4 Percobaan Skema Penilaian

Tahap selanjutnya dari proses pembuatan skema penilaian ini adalah melakukan uji coba terhadap skema penilaian yang telah disusun. Pada tahap uji coba ini, penulis meminta dua partisipan pengajar kelas menulis bahasa Indonesia untuk orang asing tingkat madya untuk mencoba melakukan penilaian tulisan menggunakan skema penilaian yang dirancang oleh penulis hasil adaptasi dari skema milik Jacobs (1981). Kedua partisipan ini bukanlah orang yang sama dalam proses pengujian skema penilaian tahap akhir nanti.

Partisipan pertama berusia 27 tahun dan memiliki pengalaman mengajar bahasa Indonesia untuk orang asing kurang lebih lima tahun. Partisipan pertama ini berjenis kelamin perempuan. Pada dua tahun pertama masa kerja, partisipan pertama ini tidak mengajar kelas menulis tingkat madya. Partisipan pertama baru mengajar kelas menulis bahasa Indonesia untuk orang asing tingkat madya pada tiga tahun terakhir.

Partisipan yang kedua juga berjenis kelamin perempuan. Pengalaman mengajar bahasa Indonesia untuk orang asing partisipan kedua selama kurang

lebih delapan tahun. Partisipan kedua ini hampir setiap tahun selama masa kerjanya minimal satu kali mengajar kelas menulis bahasa Indonesia untuk orang asing tingkat madya.

Proses pengujian tidak dilakukan secara bersamaan. Sebelum melakukan percobaan skema penilaian, penulis terlebih dahulu memberikan penjelasan mengenai jenis skema penilaian yang akan digunakan. Selain itu, penulis juga memberikan penjelasan mengenai tampilan skema penilaian serta cara pemberian angkanya.

Kedua partisipan menilai sebuah tulisan yang sama menggunakan skema penilaian yang penulis rancang. Tulisan yang dinilai berupa sebuah esai yang ditulis oleh siswa kelas menulis bahasa Indonesia untuk prang asing tingkat madya. Penulis mendampingi partisipan ketika proses penilaian berlangsung.

Setelah partisipan selesai menilai tulisan, penulis menanyakan kepada partisipan mengenai kesulitan atau kendala dalam menilai menggunakan skema penilaian ini. Para partisipan pun mengungkapkan kesulitan yang dialami dari penggunaan skema penilaian ini. Penulis mendapat masukan dari para partisipan mengenai kekurangan apa yang masih perlu diperbaiki, terutama dalam hal deskripsi naratif dari setiap kriteria penilaian.

3.5.5 Perbaikan Skema Penilaian

Tahap terakhir dari proses pembuatan skema penilaian adalah perbaikan skema penilaian. Perbaikan dilakukan dengan diskusi dan berdasarkan masukan-masukan yang didapat dari hasil uji coba. Kemudian, bentuk akhir skema penilaian ini diuji menggunakan bantuan tiga orang partisipan yang masing-masing akan menilai 36 tulisan yang sama.

Hasil akhir dari skema penilaian yang telah melalui tahapan pembuatan skema penilaian dari Mertler adalah sebagai berikut.

Tabel 3.1 Bentuk akhir skema penilaian yang diadaptasi dari Jacobs

Nama siswa:

SKOR	BOBOT	PENGUNAAN BAHASA
	22-25	Penggunaan kalimat sederhana dan kompleks yang efektif. Mahir dalam pembentukan kata menggunakan afiksasi. Sedikit kesalahan dalam penggunaan konjungsi, kata ganti, dan preposisi.
	17-21	Sedikit kesalahan dalam penggunaan kalimat sederhana dan kompleks. Kadang terjadi kesalahan afiksasi. Beberapa kesalahan dalam penggunaan konjungsi, kata ganti, dan preposisi. Makna masih jelas.
	12-16	Banyak kesalahan dalam penggunaan kalimat sederhana atau kompleks. Banyak terjadi kesalahan pemakaian afiksasi. Banyak kesalahan dalam penggunaan konjungsi, kata ganti, dan preposisi. Makna membingungkan dan tidak jelas.
	1-11	Tidak menguasai pembentukan kalimat. Sangat banyak kesalahan dalam penggunaan afiksasi, konjungsi, kata ganti, dan preposisi. Tidak bermakna.
SKOR	BOBOT	GAGASAN DAN KREATIVITAS
	12-15	Mampu mengembangkan gagasan melebihi standar. Mampu memberikan contoh-contoh, ilustrasi, data, atau fakta untuk mendukung gagasan utama.
	8-11	Mengembangkan gagasan sesuai standar yang diminta. Sedikit memberikan contoh, ilustrasi, data, atau fakta untuk mendukung gagasan utama.
	4-7	Kurang bisa mengembangkan gagasan sesuai standar. Memberikan contoh, ilustrasi, data, dan fakta, tapi tidak sesuai untuk mendukung gagasan utama.
	1-3	Sangat kurang mengembangkan gagasan. Tidak memberikan contoh, ilustrasi, data, atau fakta untuk mendukung gagasan.
SKOR	BOBOT	ORGANISASI
	18-20	Tulisan terorganisasi (ada pembuka, isi, dan penutup). Penyampaian ide secara singkat, jelas, dan runtut.
	14-17	Tulisan kurang terorganisasi, tapi pikiran utama masih dapat dimengerti. Ide disampaikan secara jelas, tapi kurang ringkas dan tidak runtut.
	10-13	Tulisan kurang terorganisasi, tapi pikiran utama tidak dapat dimengerti karena penyampaian ide yang terputus-putus dan tidak lengkap.
	1-9	Tulisan sama sekali tidak terorganisasi dan tidak runtut.
SKOR	BOBOT	KOSAKATA
	12-15	Kosakata luas. Pemakaian kata yang efektif. Penggunaan ragam (formal, informal, ragam tulis, dsb) sangat tepat.
	8-11	Kosakata cukup luas. Kadang terjadi kesalahan pemakaian kata, tapi tidak mengurangi makna tulisan. Penggunaan ragam cukup tepat.
	4-7	Kosakata terbatas. Banyak terjadi kesalahan pemakaian kata sehingga makna membingungkan dan tidak jelas. Penggunaan ragam tidak konsisten.
	1-3	Kosakata terbatas. Pemakaian kata banyak yang salah. Penggunaan ragam tidak tepat.
SKOR	BOBOT	CAPAIAN TUGAS
	12-15	Tulisan relevan dengan topik yang diminta. Tulisan sesuai dengan instruksi tugas. Wawasan pengetahuan mengenai topik yang diberikan luas.
	8-11	Tulisan mayoritas relevan dengan topik, tapi kurang detail. Tulisan mayoritas sesuai dengan instruksi tugas. Wawasan pengetahuan mengenai topik cukup luas.
	4-7	Tulisan kurang relevan dengan topik dan tidak sepenuhnya sesuai dengan instruksi. Wawasan mengenai topik terbatas.
	1-3	Tulisan tidak relevan sama sekali dengan topik dan instruksi. Tidak memiliki pengetahuan tentang topik.
SKOR	BOBOT	EJAN DAN TANDA BACA
	8-10	Penguasaan yang baik. Sedikit terjadi kesalahan dalam ejaan, tanda baca, dan huruf kapital.
	4-7	Terkadang salah dalam ejaan, tanda baca, dan huruf kapital.
	1-3	Banyak salah dalam ejaan, tanda baca, dan huruf kapital.

Total skor:

Komentar:

Pada Tabel 3.1, pada kolom pertama adalah tempat bagi penilai untuk menuliskan nilai yang akan diberikan terhadap masing-masing kriteria penilaian. Pada kolom kedua adalah bobot nilai dari setiap skala yang ada di tiap kriteria

penilaian. Sementara itu, di kolom ketiga adalah kriteria penilaian yang disusun dari atas ke bawah, begitu pula dengan skala penilaiannya.

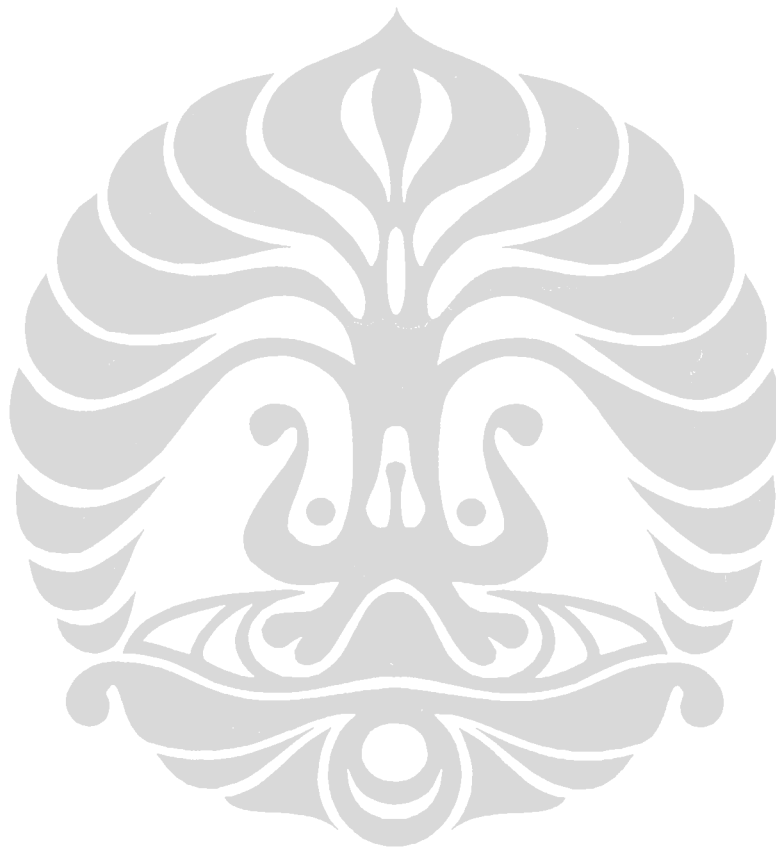
Pada bagian di bawah tabel skema penilaian, ada tempat untuk menuliskan nilai akhir yang dijumlahkan dari nilai yang diberikan pada setiap kriteria penilaian. Selain itu, ada juga bagian untuk memberi komentar tambahan mengenai tulisan yang dinilai. Tempat untuk memberi komentar tambahan itu ada di bagian paling bawah dari skema penilaian.

3.5.6 Proses Pengujian Skema Penilaian

Sebelum partisipan menguji skema penilaian yang penulis buat, mereka terlebih dahulu diberikan pelatihan mengenai pemakaian skema penilaian serta penjelasan tiap komponen yang ada di dalam skema penilaian itu. Pelatihan dilakukan secara terpisah karena ketiga partisipan memiliki kesibukan dan jadwal mengajar yang berbeda-beda. Walaupun begitu, pelatihan yang peneliti berikan memiliki isi yang sama.

Pertama-tama peneliti menerangkan mengenai tampilan skema penilaian. Peneliti meminta partisipan untuk menuliskan nama pemilik tulisan yang dinilai pada kolom di kiri atas. Kemudian, peneliti juga meminta untuk melingkari jenis tulisan yang sedang dinilai, apakah itu sampel UTS atau UAS. Kedua, peneliti menjelaskan bahwa pada skema penilaian ini ada enam kriteria penilaian yang harus diperhatikan, yaitu kriteria “Penggunaan Bahasa”, “Gagasan dan Kreativitas”, “Organisasi”, “Kosakata”, “Capaian Tugas”, dan kriteria “Ejaan dan Tanda Baca.” Peneliti juga menerangkan bahwa pada setiap kriteria penilaian itu memiliki empat level atau tingkatan kemahiran, kecuali kriteria “Ejaan dan Kosakata”, yang dicapai oleh siswa. Level-level tersebut juga memiliki deskripsi masing-masing. Peneliti juga menjelaskan mengenai deskripsi dari setiap level tersebut. Ketiga, peneliti menjelaskan bobot-bobot yang dimiliki setiap kriteria penilaian serta rentang nilai yang ada di setiap level. Peneliti menjelaskan kepada partisipan bahwa nilai yang diberikan untuk tiap kriteria dituliskan pada kolom yang ada di sebelah kiri, kemudian langkah terakhir adalah menjumlah setiap nilai yang ada di kolom paling kiri untuk mendapatkan nilai total.

Dalam pemberian pelatihan itu tidak ada batasan waktu yang peneliti tetapkan. Proses pelatihan skema penilaian itu juga tidak memakan waktu lama karena ketiga partisipan sudah cukup mengenal kriteria-kriteria penilaian yang digunakan. Setelah memberikan pelatihan itu, maka penulis berasumsi ketiga partisipan sudah memiliki keseragaman akan skema penilaian yang akan digunakan.



BAB IV

ANALISIS DAN PEMBAHASAN

Pada bab ini akan dijelaskan hasil perhitungan data secara statistik beserta pembahasannya. Hasil pengujian skema penilaian dan analisis data akan menentukan apakah skema penilaian yang telah dibuat sudah andal atau belum untuk digunakan menilai tulisan. Hasil uji statistik juga memberikan informasi keandalan antarpenilai. Hipotesis awal penelitian ini adalah jika ketiga partisipan memberikan nilai yang berbeda jauh, dapat dikatakan bahwa perbedaan mereka cukup signifikan. Sementara itu, jika ketiga partisipan memberikan nilai yang tidak berbeda jauh, dapat dikatakan tidak ada perbedaan nilai yang signifikan di antara ketiga partisipan. Skema penilaian yang penulis rancang ini dinyatakan andal apabila tidak ada perbedaan nilai yang signifikan di antara ketiga partisipan.

Sebelum melakukan analisis data, terlebih dahulu peneliti memperlihatkan soal ujian tengah semester dan soal ujian akhir semester yang menjadi dasar dari tulisan yang dihasilkan. Berikut ini adalah soal ujian tengah semester.

Pilihlah satu dari tiga pilihan di bawah ini! Kemudian, buatlah sebuah paragraf dari pilihan tersebut! Lingkari pilihan Anda! (Bobot: 50)

a. Topik : KDRT

Metode pengembangan : definisi luas

b. Topik : cara membuat mi instan

Metode pengembangan : proses

c. Topik : hal-hal yang menyebabkan banjir di Jakarta

Metode pengembangan : sebab-akibat/kausalitas

Tulisan yang dihasilkan dari soal UTS itu hanya berupa satu paragraf. Sementara itu, berikut ini adalah soal dari ujian akhir semester. Tulisan yang dihasilkan dari soal ujian akhir semester lebih panjang dari hasil tulisan UTS, yakni berupa esai.

Bacalah situasi berikut.

Anda sedang berada di ruang tunggu Bandara Soekarno-Hatta. Di sana, Anda berkenalan dengan B. Anda bertanya tujuan perjalanan B. Ternyata,

tujuan perjalanan B sama dengan tujuan perjalanan Anda. Hotel tempat B menginap juga sama dengan Anda. B bertanya kepada Anda tempat-tempat wisata yang menarik untuk dikunjungi. Anda menawarkan diri untuk menemani B berjalan-jalan di sana karena tujuan Anda ke kota itu juga untuk berwisata.

Buatlah sebuah tulisan yang di dalamnya terdapat deskripsi diri orang dan deskripsi kegiatan berdasarkan situasi di atas. Berilah judul yang menarik untuk tulisan tersebut. Ketika menulis gunakan ejaan, tanda baca, kalimat, dan kepaduan paragraf yang baik. Buat minimal 5 paragraf.

Setelah ketiga partisipan mengembalikan skema penilaian hasil menilai semua sampel tulisan, penulis memasukkan semua data ke dalam tabel supaya lebih mudah untuk dihitung secara statistik. Setiap data dari masing-masing partisipan dibagi menjadi dua, yakni data nilai UTS dan data nilai UAS. Selengkapnya dapat dilihat pada tabel-tabel berikut.

Tabel 4.1 Data Partisipan 1

Partisipan 1	
UTS	UAS
19	33
25	50
29	50
29	51
34	57
35	57
36	59
38	61
44	64
53	65
54	68
55	70
59	74
64	74
65	82
66	85
71	86
72	90

Tabel 4.2 Data Partisipan 2

Partisipan 2	
UTS	UAS
44	53
60	57
60	59
64	62
65	64
72	72
73	72
74	76
75	77
78	78
82	79
86	79
86	83
89	87
91	92
92	94
95	94
96	95

Tabel 4.3 Data Partisipan 3

Partisipan 3	
UTS	UAS
44	54
48	55
48	58
50	64
54	65
56	70
56	74
60	76
65	78
67	78
72	78
74	78
77	79
81	85
82	86
82	87
85	91
86	92

Data-data tersebut akan dianalisis secara terpisah, yakni analisis data hasil UTS dan hasil UAS. Alasannya adalah jenis tulisan yang dinilai pada UTS dan UAS berbeda. Pada UTS, tulisan yang dinilai hanya berupa satu paragraf. Sementara itu, pada UAS, tulisan yang dinilai sudah berupa esai yang terdiri dari beberapa paragraf.

Kemudian, data-data yang sudah tersusun tersebut dianalisis secara statistik. Untuk mendapatkan data statistik, dilakukan pemrosesan data hasil pengujian skema. Prosedur pemrosesan data dilakukan dalam beberapa langkah. Data hasil penghitungan statistik yang diperoleh akan digunakan untuk menilai keandalan skema penilaian.

4.1 Mean

Mean, atau nilai rata-rata, diperoleh dengan menjumlah semua angka kemudian dibagi jumlah data (n). Angka-angka tersebut berasal dari nilai yang diberikan oleh setiap partisipan terhadap tulisan yang dinilai. Berikut ini adalah tabel untuk nilai rata-rata ketiga partisipan.

Tabel 4.4 Nilai rata-rata ketiga partisipan

Partisipan 1		Partisipan 2		Partisipan 3	
UTS	UAS	UTS	UAS	UTS	UAS
64	50	75	57	72	55
65	61	96	72	81	70
53	85	74	95	67	92
35	86	73	94	56	91
55	74	78	92	74	85
29	50	44	59	44	58
29	59	60	77	48	78
54	51	86	78	77	76
66	68	95	76	86	78
19	74	65	72	54	74
72	33	89	64	82	54
44	57	82	83	65	78
25	65	60	53	50	65
36	70	72	79	60	78
71	57	86	62	82	64
34	64	64	79	48	79

	38	90	91	87	56	87
	59	82	92	94	85	86
Total	848	1176	1382	1373	1187	1348
Nilai rata-rata	47,11	65,3	76,78	76,27	65,94	74,89

Angka-angka yang ada pada Tabel 4.4 di atas merupakan nilai yang diberikan oleh masing-masing partisipan terhadap 36 tulisan. Angka-angka tersebut dibariskan berdasarkan nilai dari siswa yang sama, misalnya pada baris ketiga merupakan nilai dari siswa A dan pada baris ke-20 merupakan nilai dari siswa T. Pada tabel di atas dapat dilihat bahwa siswa A mendapat nilai UTS 64 dan nilai UAS 50 dari Partisipan 1, nilai UTS 75 dan nilai UAS 57 dari Partisipan 2, serta mendapat nilai UTS 72 dan nilai UAS 55 dari Partisipan 3. Pada baris ke-20, siswa T juga mendapat nilai 59 (UTS) dan 82 (UAS) dari Partisipan 1, nilai 92 (UTS) dan 94 (UAS) dari Partisipan 2, serta mendapat nilai 85 (UTS) dan 86 (UAS) dari Partisipan 3.

Nilai-nilai dari tiap kolom tersebut kemudian dijumlah ke bawah untuk mendapatkan jumlah total. Kemudian, jumlah total yang didapat dari tiap kolom dibagi 18 (n) untuk mendapatkan nilai rata-rata. Dari ketiga partisipan, terlihat bahwa Partisipan 1 rata-rata memberikan nilai yang cukup rendah untuk UTS dan UAS, yakni 47,11 (UTS) dan 65,3 (UAS). Sementara itu, Partisipan 2 memiliki nilai rata-rata yang tertinggi, yakni 76,78 untuk UTS dan 76,27 untuk UAS, diikuti dengan Partisipan 3 di posisi dua dengan nilai rata-rata 65,94 untuk UTS dan 74,89 untuk UAS.

4.2 Uji Normalitas

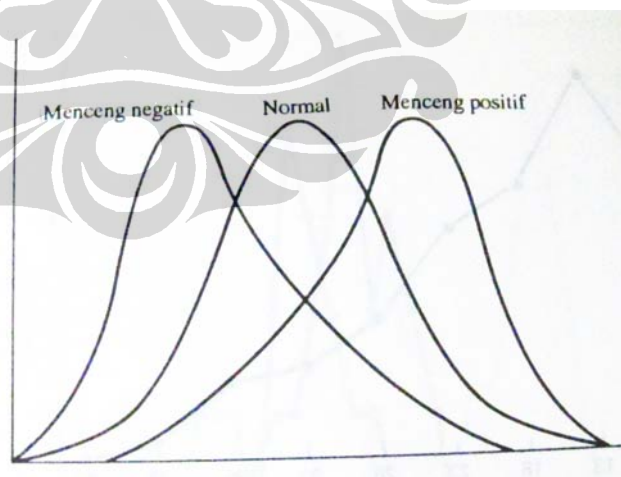
Salah satu uji statistik adalah uji normalitas data. Uji normalitas berguna untuk menentukan apakah data yang telah dikumpulkan merupakan distribusi normal atau bukan. Pengujian normalitas akan mengarahkan teknik statistik apa yang akan digunakan untuk uji pengambilan keputusan. Jika hasil uji normalitas menyatakan bahwa data telah terdistribusi normal, dapat dikatakan bahwa

kesimpulan yang didapat dari penelitian dapat digeneralisasikan ke populasi dari penelitian.

Metode statistik klasik dalam pengujian normalitas suatu data tidak begitu rumit. Berdasarkan pengalaman empiris ahli statistik, data yang banyaknya lebih dari 30 ($n > 30$) sudah dapat diasumsikan berdistribusi normal. Namun, untuk memberikan kepastian data berdistribusi normal atau tidak, sebaiknya digunakan uji normalitas, sebab belum tentu data yang lebih dari 30 bisa dipastikan berdistribusi normal, demikian juga yang kurang dari 30 belum tentu tidak berdistribusi normal, untuk itu perlu suatu pembuktian. Berikut ini beberapa cara yang umum pada pengolahan data menggunakan SPSS dalam menguji normalitas data:

1. Dengan melihat hasil nilai Skewness Kurtosis yang didapat melalui statistik deskriptif.

Uji normalitas dengan Skewness dan Kurtosis memberikan kelebihan tersendiri, yaitu bahwa akan diketahui grafik normalitas condong ke kanan, condong ke kiri, terlalu datar, atau mengumpul di tengah (lihat Gambar 4.1). Oleh karena itu, uji normalitas dengan Skewness dan Kurtosis juga sering disebut dengan ukuran kecondongan data.



Gambar 4.1 Uji normalitas Skewness Kurtosis

Kelebihan dari uji Skewness dan Kurtosis adalah bahwa kita dapat mengetahui kecondongan data. Data yang normal akan menyerupai bentuk

lonceng. Kemungkinan yang ada adalah condong ke kiri jika nilai Zskew positif dan di atas 1,96, atau condong ke kanan jika Zskew bernilai negatif dan di bawah 1,96. Syarat data yang normal adalah nilai Zskew dan Zkurt $< \pm 1,96$ (signifikansi 0,05). Untuk menghitung Zskew menggunakan rumus ($\sqrt{6/N}$) dengan N adalah jumlah observasi. Persamaan yang sama juga dipakai untuk menghitung Zkurt ($\sqrt{24/N}$).

2. Kolmogorov-Smirnov dengan pendekatan koreksi Lillifors.

Metode Lillifors menggunakan data dasar yang belum diolah dalam tabel distribusi frekuensi. Data ditransformasikan dalam nilai Z untuk dapat dihitung luasan kurva normal sebagai probabilitas kumulatif normal ($F(x)$). Probabilitas tersebut dicari bedanya dengan probabilitas kumulatif empiris ($S(x)$).

Langkah-langkah uji normalitas data dengan rumus Lillifors dilakukan sebagai berikut: Pertama, masukkan nilai atau skor pada tabel kerja secara berurutan. Kedua, mencari nilai *Z score* dengan rumus $Z = (X_i - \text{Mean})/SD$. Ketiga, menentukan nilai Z tabel $\{F(z)\}$. Keempat, menentukan $S(z)$ dengan rumus $S(z) = f \text{ kum} : N$. Kelima, menghitung harga Lillifors dengan rumus $L_h = |F(z) - S(z)|$. Keenam, mencari nilai Lillifors terbesar sebagai L_{hitung} . Ketujuh, menentukan harga Lillifors tabel (L_t) dengan rumus (a, n). Kedelapan, membuat kesimpulan: Data berdistribusi normal jika harga $L_h < t$, sedangkan data tidak berdistribusi normal jika harga $L_h > \text{harga } L_t$.

3. Kolmogorov-Smirnov satu-sampel.

Konsep dasar dari uji normalitas Kolmogorov Smirnov adalah dengan membandingkan distribusi data (yang akan diuji normalitasnya) dengan distribusi normal baku. Distribusi normal baku adalah data yang telah ditransformasikan ke dalam bentuk Z-Score dan diasumsikan normal. Jadi, sebenarnya uji Kolmogorov Smirnov adalah uji beda antara data yang diuji normalitasnya dengan data normal baku. Seperti pada uji beda biasa, jika

signifikansi di bawah 0,05 berarti terdapat perbedaan yang signifikan, dan jika signifikansi di atas 0,05 berarti tidak terjadi perbedaan yang signifikan.

Penelitian ini menggunakan tes Kolmogorov-Smirnov satu-sampel untuk menguji normalitas data. Tes Kolmogorov-Smirnov satu-sampel ini bertujuan melihat apakah di antara dua set data memiliki perbedaan yang signifikan. Pertama adalah uji normalitas data UTS. Jumlah data (n) UTS dari ketiga partisipan ada 54. Uji normalitas data UTS dapat dilihat pada Tabel 4.5.

Tabel 4.5 Uji normalitas UTS

One-Sample Kolmogorov-Smirnov Test		UTS
N		54
Normal Parameters ^{a,b}	Mean	63.2778
	Std. Deviation	19.41641
Most Extreme Differences	Absolute	.081
	Positive	.052
	Negative	-.081
Kolmogorov-Smirnov Z		.594
Asymp. Sig. (2-tailed)		.873

a. Distribusi tes adalah normal.

b. Dikalkulasi berdasarkan data.

Tabel 4.5 menunjukkan hasil uji normalitas untuk data UTS. Jika nilai signifikansi lebih besar dari 0,05, data dinyatakan tidak memiliki perbedaan yang signifikan. Sementara itu, jika nilai uji berkisar antara 0 sampai 0,05, data bisa dikatakan memiliki perbedaan yang signifikan. Dari Tabel 4.5 kita melihat hasil yang diperoleh menunjukkan nilai signifikansi sebesar 0,873. Nilai ini lebih besar dari 0,05 sehingga dapat disimpulkan bahwa data UTS sudah mengikuti sebaran normal.

Tabel 4.6 di bawah ini menunjukkan uji normalitas data UAS menggunakan tes Kolmogorov-Smirnov satu-sampel. Jumlah data UAS juga sama seperti data UAS, yakni 54 data dari tiga partisipan.

Tabel 4.6 Uji normalitas UAS

		UAS
N		54
Normal Parameters ^{a,b}	Mean	72.1667
	Std. Deviation	14.02525
Most Extreme Differences	Absolute	.089
	Positive	.067
	Negative	-.089
Kolmogorov-Smirnov Z		.655
Asymp. Sig. (2-tailed)		.784

a. Distribusi tes adalah normal.

b. Dikalkulasi berdasarkan data.

Dari Tabel 4.6 terlihat hasil uji normalitas untuk data UAS. Hasil yang diperoleh menunjukkan nilai signifikansi sebesar 0,784. Nilai ini lebih besar dari 0,05 sehingga dapat disimpulkan bahwa data UAS juga sudah mengikuti sebaran normal.

4.3 Uji Homogenitas Varian

Uji homogenitas varian berfungsi untuk memberikan keyakinan bahwa sekumpulan data yang akan diuji dalam serangkaian analisis memang berasal dari populasi yang tidak jauh berbeda keragamannya (homogen). Uji homogenitas varian dapat dilakukan apabila kelompok data tersebut memiliki distribusi normal.

Apabila hasil uji homogenitas varian tidak homogen, pengolahan data tidak dapat dilanjutkan ke dalam pengukuran dan pengujian hipotesis. Alasannya, data yang didapatkan dari para responden atau partisipan tidak merepresentasikan keseluruhan responden atau partisipan secara benar menurut keadaan yang sebenarnya. Uji homogenitas varian yang pertama adalah data UTS, yang selengkapnya dapat dilihat pada Tabel 4.7.

Tabel 4.7 Uji homogenitas varian UTS

UTS			
Levene Statistic	df1	df2	Sig.
1.118	2	51	.335

Tabel di atas menunjukkan hasil uji homogenitas varian untuk data UTS. Hasil yang diperoleh menunjukkan nilai signifikansi sebesar 0,335. Nilai ini lebih besar dari 0,05 sehingga dapat disimpulkan bahwa data UTS memiliki varian yang homogen.

Uji homogenitas yang kedua dilakukan terhadap data UAS, yang selengkapnya dapat dilihat pada Tabel 4.8 di bawah ini.

Tabel 4.8 Uji homogenitas varian UAS

UAS			
Levene Statistic	df1	df2	Sig.
.455	2	51	.637

Dari Tabel 4.8 di atas kita dapat melihat hasil uji homogenitas varian untuk data UAS. Hasil yang diperoleh menunjukkan nilai signifikansi sebesar 0,637. Nilai ini lebih besar dari 0,05 sehingga dapat disimpulkan bahwa data UAS memiliki varian yang homogen.

4.4 Uji Perbandingan

Analisis perbandingan digunakan untuk membandingkan rata-rata antara dua atau lebih kelompok sampel data. Asumsi mendasar dalam analisis perbandingan adalah bahwa variabel data yang akan dibandingkan harus mengikuti distribusi normal. Asumsi lainnya yang harus dipenuhi dalam analisis perbandingan dengan ANOVA (*Analysis of Variance*) adalah homogenitas varian. Ini dilakukan melalui uji *Levene's homogeneity-of-variance test*.

Data dalam penelitian ini sudah melalui proses pengujian distribusi dan uji homogenitas varian. Hasil dari kedua pengujian itu menyatakan bahwa data penelitian ini sudah terdistribusi normal dan memiliki varian yang homogen. Oleh karena itu, pengujian data ini dapat dilanjutkan ke tahap selanjutnya, yaitu uji perbandingan.

Uji perbandingan ada dua macam, yaitu uji t dan analisis varians (ANOVA). Statistik uji t dan ANOVA digunakan sebagai statistik uji untuk

perbandingan dua atau lebih kelompok sampel data. Uji t digunakan untuk membandingkan dua sampel yang akan dibandingkan, sedangkan ANOVA digunakan untuk uji perbandingan lebih dari dua kelompok sampel data. Data yang digunakan dalam penelitian ini terbagi dalam tiga kelompok, karena itu akan diuji menggunakan ANOVA. Uji perbandingan yang pertama adalah uji data UTS.

Tabel 4.9 Deskriptif UTS

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Partisipan 1	18	47.1111	16.99096	4.00481	38.6617	55.5605	19.00	72.00
Partisipan 2	18	76.7778	14.28926	3.36801	69.6719	83.8837	44.00	96.00
Partisipan 3	18	65.9444	14.32592	3.37665	58.8203	73.0686	44.00	86.00
Total	54	63.2778	19.41641	2.64224	57.9781	68.5774	19.00	96.00

Dari Tabel 4.9 dapat dilihat perbedaan deskriptif dari data UTS. Partisipan 1 memiliki nilai rata-rata 47,1111 dengan *standard deviation* 16,99096 dan *standard error* 4,00481. Partisipan 2 memiliki nilai rata-rata 76,7778 dengan *standard deviation* 14,28926 dan *standard error* 3,36801. Partisipan 3 memiliki nilai rata-rata 65,9444 dengan *standard deviation* 14,32592 dan *standard error* 3,37665.

Tabel 4.10 Uji perbandingan UTS

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	8113.000	2	4056.500	17.432	.000
Within Groups	11867.833	51	232.703		
Total	19980.833	53			

Tabel ANOVA di atas menunjukkan hasil uji perbandingan nilai UTS atas tiga orang penguji. Nilai tertinggi diberikan oleh Partisipan 2 dengan nilai 76,78 diikuti Partisipan 3 dengan 65,94, dan paling rendah Partisipan 1 sebesar 47,11 dengan nilai rata-rata keseluruhan sebesar 63,78.

Nilai signifikansi yang diperoleh adalah 0,000. Nilai tersebut lebih kecil dari 0,05 sehingga dapat disimpulkan bahwa ketiga nilai tersebut memiliki perbedaan yang signifikan. Weigle (2002) mengatakan jika nilai mendekati angka 0 berarti tidak ada atau hanya sedikit korelasi yang terjadi antara nilai yang diberikan oleh para penilai. Sementara itu, jika nilainya mendekati angka 1 berarti ada korelasi yang kuat di antara para penilai.

Untuk mengetahui penilai mana yang lebih kecil atau lebih besar dan memiliki signifikansi dalam perbedaannya, dilakukan pengujian lanjutan dengan uji Duncan. Uji Duncan adalah uji lanjutan untuk mengetahui nilai rata-rata mana saja yang sama dan nilai rata-rata mana saja yang tidak sama ketika pengujian homogenitas beberapa nilai rata-rata memberikan hasil menolak hipotesis nol dan menerima hipotesis alternatif.

Tabel 4.11 Uji Duncan UTS

Duncan ^a		Subset for alpha = .05		
Penguji	N	1	2	3
Partisipan 1	18	47.1111		
Partisipan 3	18		65.9444	
Partisipan 2	18			76.7778
Sig.		1.000	1.000	1.000

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 18.000.

Hasil uji Duncan menunjukkan bahwa perbedaan antara ketiga penilai tersebut signifikan secara statistik. Hal ini ditunjukkan oleh nilai yang saling terpisah dalam masing-masing kolom sebagaimana yang ditampilkan pada Tabel 4.11 di atas. Berdasarkan hasil ini dapat disimpulkan bahwa secara statistik Partisipan 1 memiliki nilai terkecil diikuti oleh Partisipan 3 dan yang tertinggi dimiliki oleh Partisipan 2.

Uji perbandingan yang kedua adalah uji data UAS.

Tabel 4.12 Deskriptif UAS

UAS	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
					Partisipan 1	18		
Partisipan 2	18	76.2778	13.30106	3.13509	69.6633	82.8922	53.00	95.00
Partisipan 3	18	74.8889	11.73175	2.76520	69.0548	80.7230	54.00	92.00
Total	54	72.1667	14.02525	1.90859	68.3385	75.9948	33.00	95.00

Dari Tabel 4.12 di atas dapat dilihat penjabaran deskriptif data UAS. Partisipan 1 memiliki nilai rata-rata 65,3333 dengan *standard deviation* 14,95090 dan *standard error* 3,52396. Partisipan 2 memiliki nilai rata-rata 76,2778 dengan *standard deviation* 13,30106 dan *standard error* 3,13509. Partisipan 3 memiliki nilai rata-rata 74,8889 dengan *standard deviation* 11,73175 dan *standard error* 2,76520. Nilai terendah diberikan oleh Partisipan 1, yakni 33, sedangkan nilai tertinggi diberikan oleh Partisipan 2, yakni 95.

Tabel 4.13 Uji perbandingan UAS

UAS	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1278.111	2	639.056	3.563	.036
Within Groups	9147.389	51	179.361		
Total	10425.500	53			

Tabel ANOVA di atas menunjukkan hasil uji perbandingan nilai UAS atas tiga orang penguji. Nilai tertinggi diberikan oleh Partisipan 2 dengan nilai 76,28 diikuti Partisipan 3 dengan 74,89 dan paling rendah Partisipan 1 sebesar 65,33 dengan nilai rata-rata keseluruhan sebesar 72,17.

Nilai signifikansi yang diperoleh adalah sebesar 0,036. Nilai tersebut lebih kecil dari 0,05 sehingga dapat disimpulkan bahwa ketiga nilai tersebut memiliki perbedaan yang signifikan. Untuk mengetahui penilai mana yang lebih kecil atau lebih besar dan memiliki signifikansi dalam perbedaannya, dilakukan pengujian lanjutan dengan uji Duncan.

Tabel 4.14 Uji Duncan UAS

Duncan^a

Penguji	N	Subset for alpha = .05	
		1	2
Partisipan 1	18	65.3333	
Partisipan 3	18		74.8889
Partisipan 2	18		76.2778
Sig.		1.000	.757

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 18.000.

Hasil uji Duncan menunjukkan bahwa perbedaan antara ketiga penilai tersebut membentuk dua kelompok yang berbeda signifikan secara statistik. Kelompok pertama hanya ditempati oleh Partisipan 1, yaitu dengan nilai yang lebih rendah, dan kelompok yang lebih besar ditempati oleh Partisipan 3 dan Partisipan 2 yang tidak berbeda secara statistik.

Berdasarkan hasil ini dapat disimpulkan bahwa secara statistik Partisipan 1 memiliki nilai terkecil diikuti oleh Partisipan 3 dan Partisipan 2 yang tergabung dalam satu kelompok. Selain itu, dari keseluruhan analisis data, terlihat bahwa Partisipan 1 selalu memberikan nilai terendah, yakni 19 pada UTS dan 33 pada UAS. Sementara itu, Partisipan 2 selalu memberikan nilai tertinggi, yakni 96 pada UTS dan 95 pada UAS.

4.5 Pembahasan

Setelah menganalisis data hasil penilaian secara statistik, penulis akan membahas dua hal penting terkait skema penilaian ini. Pertama berkaitan dengan keandalan antarpenilai dalam menilai tulisan, baik yang hanya satu paragraf maupun tulisan esai. Kedua berkaitan dengan skema penilaian yang penulis buat.

Hasil uji statistik data UTS memperlihatkan bahwa di antara ketiga partisipan tidak memiliki keseragaman nilai. Hal itu dapat dilihat pada Tabel 4.10 dan Tabel 4.11. Nilai signifikansi yang dihasilkan dari penghitungan statistik menunjukkan angka 0,000. Nilai tersebut lebih kecil dari 0,05, dengan kata lain ketiga penilai tidak memiliki keandalan antarpenilai dalam menilai tulisan UTS yang berbentuk paragraf. Selain itu, dari Tabel 4.11 juga terlihat bahwa ketiga

partisipasi berada dalam kelompok nilai yang berbeda. Oleh karena itu, penulis menyimpulkan bahwa skema penilaian ini tidak cocok untuk menilai sebuah paragraf. Hal itu terlihat dari perbedaan nilai di antara ketiga penilai yang cukup signifikan berdasarkan uji ANOVA. Namun, skema penilaian ini masih dapat diandalkan untuk menilai sebuah esai. Dari uji ANOVA terhadap data UAS, terlihat bahwa hanya Partisipan 1 yang memiliki nilai beda, sedangkan Partisipan 2 dan Partisipan 3 termasuk ke dalam kelompok yang sama. Perbedaan itu dapat disebabkan interpretasi yang berbeda oleh Partisipan 1 terhadap tulisan sehingga ia memberikan nilai yang lebih rendah dibandingkan nilai yang diberikan oleh Partisipan 2 dan Partisipan 3. Jadi, dapat disimpulkan bahwa antara Partisipan 2 dan Partisipan 3 sudah memiliki keandalan antarpemilai. Dengan kata lain, Partisipan 2 dan Partisipan 3 sudah memiliki kesamaan pandangan terhadap skema penilaian yang penulis buat.

Selain itu, dari hasil uji statistik data UTS dan UAS, dapat disimpulkan juga bahwa Partisipan 1 selalu masuk dalam kelompok yang memiliki nilai terkecil. Sementara itu, Partisipan 2 selalu masuk dalam kelompok dengan nilai terbesar. Hal itu dapat dilihat pada Tabel 4.11 dan Tabel 4.14. Partisipan 1 juga cenderung memberikan nilai yang rendah terhadap tulisan siswa. Hal itu dapat dilihat dari nilai rata-rata pada data UTS dan data UAS Partisipan 1. Pada UTS, Partisipan 1 memberikan nilai rata-rata 47,1111, sedangkan pada UAS Partisipan 1 memberikan nilai rata-rata 65,3333. Nilai rata-rata itu yang terendah jika dibandingkan dengan nilai rata-rata yang dimiliki oleh Partisipan 2 dan Partisipan 3. Partisipan 2 memberikan nilai rata-rata 76,7778 untuk UTS dan nilai rata-rata 76,2778 untuk UAS. Sementara itu, Partisipan 3 memberikan nilai rata-rata 65,9444 untuk UTS dan nilai rata-rata 74,8889 untuk UAS.

Dari pembahasan di atas terlihat bahwa skema penilaian ini belum cukup andal digunakan untuk menilai tulisan yang hanya terdiri dari satu paragraf. Hal itu dapat dilihat dari Tabel 4.11 yang memperlihatkan bahwa ketiga penilai tidak memiliki keandalan antarpemilai yang sama. Namun, skema penilaian ini masih cukup andal digunakan untuk menilai tulisan berupa esai. Hal itu terlihat pada

Tabel 4.14 yang memperlihatkan bahwa dua dari tiga penilai memiliki keandalan yang sama.

Melihat adanya perbedaan yang cukup signifikan, terutama bagi Partisipan 1 dalam penilaian UTS, penulis mengadakan proses lanjutan dari penelitian ini, yakni wawancara kepada partisipan. Wawancara ini untuk mengetahui alasan kenapa Partisipan 1 memberikan nilai yang cukup rendah dibandingkan kedua partisipan lainnya. Pertama-tama penulis menanyakan kepada para partisipan mengenai perbedaan cara menilai tulisan esai dan tulisan satu paragraf menggunakan skema penilaian ini. Di antara ketiga partisipan, hanya Partisipan 1 yang merasa kesulitan menggunakan skema penilaian ini untuk menilai sebuah paragraf. Ketika diwawancara, Partisipan 1 menjelaskan alasan ia memberikan nilai yang rata-rata rendah terhadap tulisan-tulisan hasil UTS. Menurutnya, level-level yang ada di skema penilaian ini sangat tinggi kriterianya sehingga kurang cocok untuk diterapkan ke sebuah paragraf. Skema penilaian ini lebih cocok digunakan untuk menilai esai yang panjang. Menurut Partisipan 1, panjang tulisan sangat memengaruhi penilaian. Alasannya adalah jika ada dua atau tiga kesalahan tanda baca dalam sebuah paragraf, hal itu menjadi terlihat menonjol karena tulisannya hanya sebuah paragraf. Lain halnya jika kesalahan itu muncul pada sebuah esai, maka kesalahan itu tidak akan terlihat begitu jelas. Penulis menangkap hal itulah yang menjadikan dasar kenapa nilai yang diberikan Partisipan 1 terhadap tulisan-tulisan hasil UTS sangat rendah dibandingkan yang lain. Kedua partisipan yang lain memang juga menganggap level yang ada dalam skema penilaian ini agak sulit untuk diterapkan ke sebuah paragraf. Namun, mereka cukup mahir untuk mengira-ngira persentase yang harus digunakan untuk menilai sebuah paragraf. Setelah diusut, ternyata penulis mengetahui bahwa jam terbang Partisipan 1 dalam mengajar kelas menulis bahasa Indonesia untuk orang asing tingkat madya paling sedikit di antara ketiga partisipan. Selain itu, dari hasil wawancara pun Partisipan 1 mengakui bahwa ia memiliki standar yang cukup tinggi dalam memberikan nilai. Misalnya saja jika ada satu murid yang menghasilkan tulisan sangat pendek, sudah tentu mendapat nilai yang sangat kurang. Alasannya adalah murid-murid yang lain dapat membuat tulisan yang

agak panjang, lalu kenapa murid itu tidak mampu membuat tulisan yang panjang juga.

Yang kedua penulis menanyakan kepada ketiga partisipan mengenai kesulitan dan kemudahan menggunakan skema penilaian yang digunakan. Ketiga partisipan memberikan jawaban yang tidak jauh berbeda, yakni mereka pada awalnya agak kebingungan membaca deskripsi yang panjang-panjang dari tiap kriteria. Kebingungan itu karena mereka belum hafal akan deskripsi dari setiap kriteria jadi harus bolak-balik melihat skema penilaian setiap kali akan menilai. Akan tetapi, hal itu hanya berlangsung pada awal-awal. Setelah beberapa kali menilai, mereka mulai terbiasa dengan skema penilaian. Hal itu terkait dengan kelebihan yang dimiliki skema penilaian ini, yakni mudah digunakan. Ketiga partisipan memberikan tanggapan yang positif terhadap skema penilaian ini. Menurut mereka, dengan adanya pembagian level dan deskripsi dari tiap level, memudahkan mereka untuk menentukan kemahiran murid termasuk level yang mana. Selain itu, adanya rentang nilai juga sangat membantu partisipan ketika misalnya ada satu kriteria yang menurutnya cukup bagus, tapi belum pantas juga mendapat nilai sempurna. Karena itu, partisipan memberikan nilai yang ada di tengah-tengah rentang nilai yang ada.

Partisipan 1 juga menambahkan satu kendala lain, yakni penggunaan kata *sering* dan *jarang*. Menurutnya, seberapa sering atau seberapa jarang seorang murid melakukan kesalahan itu sangat sulit untuk ditentukan. Ia kesulitan untuk menentukan apakah kesalahan yang dilakukan itu sering atau tidak. Karena itu, ia menyarankan agar kata *sering* dan *jarang* diganti saja dengan kata *banyak* dan *sedikit*, sebab ia merasa lebih mudah mengira-mengira apabila kesalahan yang dilakukan itu termasuk banyak atau sedikit.

Secara tampilan, skema penilaian yang penulis buat sudah cukup praktis digunakan, sebab skema penilaian ini hanya satu lembar kertas yang sudah memuat segala informasi untuk digunakan menilai tulisan. Seperti yang Brown (2004) katakan bahwa suatu penilaian dinyatakan praktis apabila memenuhi persyaratan tertentu, yakni tidak mahal, efisien, dan mudah digunakan. Selain itu, dari hasil wawancara pun para partisipan menyatakan bahwa skema penilaian ini

mudah digunakan dan juga mempermudah proses penilaian. Walaupun pada awalnya para penilai agak bingung menggunakan skema penilaian ini karena deskripsi yang panjang, tetapi pada akhirnya mereka merasa skema penilaian ini cukup mudah digunakan. Hal itu menjadi masukan bagi penulis untuk memberikan pelatihan yang lebih intensif kepada para penilai yang akan menggunakan skema penilaian ini agar tidak kebingungan menggunakannya.

4.6 Tindak Lanjut

Hasil dari wawancara kepada tiga partisipan akan penulis jadikan dasar perbaikan bagi skema penilaian ini. Akan tetapi, karena keterbatasan waktu, skema yang telah diperbaiki ini tidak diujikan lagi untuk melihat apakah sudah layak digunakan secara nyata atau belum.

Perbaikan pertama penulis lakukan terhadap penggunaan kata *sering* dan *jarang* dalam skema penilaian ini. Untuk menghindari kebingungan, kedua kata itu penulis ganti dengan kata *banyak* dan *sedikit*. Berdasarkan hasil wawancara, partisipan mengatakan lebih mudah mengira-ngira menggunakan kata *banyak* daripada menggunakan kata *sering*. Mereka menggunakan persentase di atas 70 persen termasuk kategori banyak, sedangkan di bawah 40 persen termasuk kategori sedikit, dan antara 40-70 persen termasuk kategori sedang. Hal itu berkaitan dengan perbaikan yang kedua, yaitu penulis menambahkan informasi tersebut di bagian bawah skema penilaian.

Selain itu, perbaikan yang terakhir adalah dalam hal pemberian penjelasan atau pelatihan kepada para penilai yang akan menggunakan skema penilaian ini. Walaupun hal tersebut ini belum penulis uji cobakan, tetapi penulis berharap ke depannya para penilai yang akan menggunakan skema penilaian harus mendapat pelatihan yang intensif mengenai penggunaan skema penilaian yang akan digunakan. Berikut ini adalah skema penilaian yang telah diperbaiki.

Tabel 4.15 Skema penilaian adaptasi dari Jacobs yang telah diperbaiki

Nama siswa:

SKOR	BOBOT	PENGGUNAAN BAHASA
	22-25	Penggunaan kalimat sederhana dan kompleks yang efektif. Mahir dalam pembentukan kata menggunakan afiksasi. Sedikit kesalahan dalam penggunaan konjungsi, kata ganti, dan preposisi.
	17-21	Sedikit kesalahan dalam penggunaan kalimat sederhana dan kompleks. Kadang terjadi kesalahan afiksasi. Beberapa kesalahan dalam penggunaan konjungsi, kata ganti, dan preposisi. Makna masih jelas.
	12-16	Banyak kesalahan dalam penggunaan kalimat sederhana atau kompleks. Banyak terjadi kesalahan pemakaian afiksasi. Banyak kesalahan dalam penggunaan konjungsi, kata ganti, dan preposisi. Makna membingungkan dan tidak jelas.
	1-11	Tidak menguasai pembentukan kalimat. Sangat banyak kesalahan dalam penggunaan afiksasi, konjungsi, kata ganti, dan preposisi. Tidak bermakna.
SKOR	BOBOT	GAGASAN DAN KREATIVITAS
	12-15	Mampu mengembangkan gagasan melebihi standar. Mampu memberikan contoh-contoh, ilustrasi, data, atau fakta untuk mendukung gagasan utama.
	8-11	Mengembangkan gagasan sesuai standar yang diminta. Sedikit memberikan contoh, ilustrasi, data, atau fakta untuk mendukung gagasan utama.
	4-7	Kurang bisa mengembangkan gagasan sesuai standar. Memberikan contoh, ilustrasi, data, dan fakta, tapi tidak sesuai untuk mendukung gagasan utama.
	1-3	Sangat kurang mengembangkan gagasan. Tidak memberikan contoh, ilustrasi, data, atau fakta untuk mendukung gagasan.
SKOR	BOBOT	ORGANISASI
	18-20	Tulisan terorganisasi (ada pembuka, isi, dan penutup). Penyampaian ide secara singkat, jelas, dan runtut.
	14-17	Tulisan kurang terorganisasi, tapi pikiran utama masih dapat dimengerti. Ide disampaikan secara jelas, tapi kurang ringkas dan tidak runtut.
	10-13	Tulisan kurang terorganisasi, tapi pikiran utama tidak dapat dimengerti karena penyampaian ide yang terputus-putus dan tidak lengkap.
	1-9	Tulisan sama sekali tidak terorganisasi dan tidak runtut.
SKOR	BOBOT	KOSAKATA
	12-15	Kosakata luas. Pemakaian kata yang efektif. Penggunaan ragam (formal, informal, ragam tulis, dsb) sangat tepat.
	8-11	Kosakata cukup luas. Kadang terjadi kesalahan pemakaian kata, tapi tidak mengurangi makna tulisan. Penggunaan ragam cukup tepat.
	4-7	Kosakata terbatas. Banyak terjadi kesalahan pemakaian kata sehingga makna membingungkan dan tidak jelas. Penggunaan ragam tidak konsisten.
	1-3	Kosakata terbatas. Pemakaian kata banyak yang salah. Penggunaan ragam tidak tepat.
SKOR	BOBOT	CAPAIAN TUGAS
	12-15	Tulisan relevan dengan topik yang diminta. Tulisan sesuai dengan instruksi tugas. Wawasan pengetahuan mengenai topik yang diberikan luas.
	8-11	Tulisan mayoritas relevan dengan topik, tapi kurang detail. Tulisan mayoritas sesuai dengan instruksi tugas. Wawasan pengetahuan mengenai topik cukup luas.
	4-7	Tulisan kurang relevan dengan topik dan tidak sepenuhnya sesuai dengan instruksi. Wawasan mengenai topik terbatas.
	1-3	Tulisan tidak relevan sama sekali dengan topik dan instruksi. Tidak memiliki pengetahuan tentang topik.
SKOR	BOBOT	EJAN DAN TANDA BACA
	8-10	Penguasaan yang baik. Sedikit terjadi kesalahan dalam ejaan, tanda baca, dan huruf kapital.
	4-7	Terkadang salah dalam ejaan, tanda baca, dan huruf kapital.
	1-3	Banyak salah dalam ejaan, tanda baca, dan huruf kapital.

Total skor:

Catatan
 Sedikit: < 40%
 Sedang/kadang: 40-70%
 Banyak: > 70%

Komentar:

BAB 5

PENUTUP

Pada bab ini akan diuraikan simpulan penelitian dan implikasinya dalam pengajaran bahasa, serta ditutup dengan pemberian saran untuk penelitian selanjutnya.

5.1 Simpulan

Tujuan penelitian ini adalah membuat skema penilaian yang dapat digunakan untuk menilai tugas menulis siswa di lembaga pengajaran bahasa Indonesia untuk orang asing dengan mudah dan andal. Melihat tujuan penelitian ini, penulis dapat menyimpulkan bahwa skema penilaian yang dihasilkan dapat dengan mudah digunakan menilai tulisan siswa. Dari hasil wawancara pun ketiga partisipan menyatakan bahwa skema penilaian ini mudah digunakan dan juga mempermudah proses penilaian. Implikasinya adalah proses penilaian pun jadi lebih cepat. Hal itu berarti langkah-langkah pembuatan skema penilaian yang penulis lakukan sudah sangat tepat.

Akan tetapi, ternyata hasil dari uji skema penilaian ini belum mencapai tingkat keandalan yang 100 persen. Dari hasil analisis terlihat bahwa tidak terdapat keandalan antarpemilai dalam menggunakan skema penilaian ini. Para pemilai memiliki perbedaan yang signifikan dalam memberikan nilai, terutama untuk tulisan hasil UTS. Dari hasil wawancara lanjutan, penulis pun mengetahui bahwa level-level yang ada dalam skema penilaian ini kurang cocok untuk digunakan menilai sebuah paragraf. Dari hasil wawancara itu juga penulis mengadakan perbaikan skema penilaian ini.

Kesimpulan kedua yang dapat ditarik dari hasil analisis adalah skema penilaian yang penulis rancang belum menunjukkan keandalan antarpemilai yang tinggi. Jika dilihat dari hasil wawancara, perbedaan nilai yang diberikan oleh ketiga partisipan bukan karena proses pembuatan skema penilaian yang salah, melainkan karena proses pelatihan penggunaan skema penilaian yang kurang intensif.

Walaupun para partisipan dapat menggunakan skema penilaian ini untuk menilai sebuah paragraf, tetapi ternyata mereka mengerjakannya lebih lambat dibandingkan ketika menilai sebuah esai. Alasannya karena ada beberapa kriteria pada skema penilaian ini yang agak sulit untuk diterapkan ke sebuah paragraf, karena itu para partisipan harus menyesuaikan sendiri kriteria-kriteria itu ketika harus menilai sebuah paragraf. Akibatnya adalah nilai yang diberikan ketiga partisipan memiliki perbedaan yang sangat signifikan. Oleh karena itu, penulis menyimpulkan para penilai perlu diberikan pelatihan yang lebih detail lagi jika ingin menggunakan skema penilaian ini untuk menilai paragraf.

5.2 Saran

Prosedur pembuatan skema penilaian yang efektif tidak diragukan lagi sangat penting dilakukan. Hal tersebut menuntut waktu dan usaha yang cukup banyak dari pembuat skema penelitian dan para partisipan. Penyusunan skema penilaian tidak dapat dilakukan secara terburu-buru. Selain itu, dalam penyusunan skema penilaian juga dibutuhkan kerja sama yang sangat baik antara peneliti, responden, dan pengajar kelas.

Saran yang kedua adalah penelitian dalam bidang ini harus dilakukan karena hasilnya sangat berguna bagi segala pihak yang terkait dengan penelitian ini. Selain membuat skema penilaian, penulis juga menyarankan kepada para pengajar atau lembaga pengajaran untuk memberikan pelatihan dan praktik yang mendalam bagi para penilai yang akan menggunakan skema penilaian, sebab hal itu akan sangat berpengaruh terhadap nilai yang diberikan dan keandalan interpenilai itu sendiri. Seperti yang Lovorn dan Rezaei (2011) katakan, penggunaan skema penilaian yang salah lebih buruk hasilnya daripada tidak menggunakan skema penilaian sama sekali.

Skema penilaian yang penulis rancang ini sangat baik untuk memberikan balikan, baik balikan kepada siswa maupun pengajar. Hal tersebut terlihat dari kriteria-kriteria yang ada dalam skema penilaian serta level yang ada di tiap kriteria. Bagi murid, balikan ini berfungsi mengetahui di mana letak kekurangan serta kelebihan mereka. Dengan mengetahui letak kekurangannya, mereka

menjadi lebih fokus untuk meningkatkan performa mereka di kekurangan itu. Sementara, bagi pengajar, balikan ini berfungsi untuk menyusun materi pembelajaran selanjutnya. Jika pengajar melihat bahwa mayoritas murid tidak mengalami perkembangan di satu kriteria, tentu pengajar harus memasukkan kriteria itu ke dalam materi pembelajaran berikutnya. Oleh karena itu, saran penulis bagi para guru adalah untuk memberikan skema penilaian ini sebelum tes berlangsung agar para murid mengetahui kriteria-kriteria yang harus dipenuhi.

Sebagai penutup, penulis berpendapat bahwa penelitian ini masih belum selesai. Masih banyak hal yang harus diperbaiki dalam skema penilaian ini untuk mencapai keandalan yang 100 persen. Penulis menyadari bahwa masih banyak kekurangan yang terdapat dalam penelitian ini. Oleh karena itu, kritik dan saran sangat penulis harapkan guna menyempurnakan kekurangan yang ada. Selain itu, penulis berharap akan ada peneliti lain yang dapat melanjutkan penelitian serupa demi perkembangan dan masukan bagi dunia pengajaran bahasa, terutama pengajaran bahasa Indonesia untuk orang asing. Dari penelitian ini pun diharapkan dapat memancing peneliti lain untuk membuat skema penilaian di bidang studi lain dan pada tingkatan yang lain.

DAFTAR REFERENSI

- Allison, Desmond M. (1999). *Language testing and evaluation: An Introductory Course*. Singapore: Singapore University Press.
- Allwright, D. dan K.M. Bailey. (1991). *Focus on the classroom: An introduction to classroom research for language teachers*. Cambridge: Cambridge University Press.
- Andrade, H.G. (1997). "Understanding rubrics". *Educational Leadership*, 54, 4. 23 November 2008.
- Bachman, Lyle F. dan A.S. Palmer. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Brown, H. Douglas. (2004). *Language assessment: Principles and classroom practices*. London: Longman.
- Brown, H. Douglas. (2007). *Principles of language learning and teaching*. Cambridge: Cambridge University Press.
- Blok, Henk dan Kees de Glopper. (1992). "Large scale writing assessment". *The construct of language proficiency: Applications of psychological models to language assessment*. Ludo Verhoven dan John H.A.L. De Jong (ed.).
- Broad, Bob. (2003). *What we really value: Beyond rubrics in teaching and assessing writing*. Utah: Utah State University Press.
- Butler, Christopher. (1985). *Statistics in linguistics* (Terj.). Bandung: ITB.
- Coombe, Christine. (2010). "Assessing foreign/second language writing ability." ProQuest.
- Cooper, C. (1977). "Holistic evaluation of writing". *Evaluating writing: Describing, measuring, judging*. C. Cooper dan L. Idell (ed.). Urbana, IL: National Council of Teachers of English.
- Council of Europe. (2001). *Common european framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Davies, Alan. (1999). *Dictionary of language testing*. Cambridge: Cambridge University Press.

- Goodrich, Heidi. (1997). "Understanding rubrics". ProQuest.
- Hadley, A. (1993). *Teaching language in context*. Boston, MA: Heinle & Heinle Publishers.
- Haines, Catherine. (2004). *Assessing students written work: Marking essays and reports*. London and New York: RoutledgeFalmer.
- Hamp-Lyons, L. (1990). "Second language writing: Assessment issues". *Second language writing: Research insight for the classroom*. Barbara Kroll (ed.). Cambridge: Cambridge University Press.
- Hamp-Lyons, L. (1991a). "Basic concepts". *Assessing second language writing in academic contexts*. L. Hamp-Lyons (ed.). Norwood, NJ: Ablex.
- Harmer, Jeremy. (2007). *The practice of english language teaching 4th edition*. London: Longman.
- Hidayat, Rahayu S. (1990). *Pengetesan kemampuan membaca secara komunikatif*. Jakarta: Linimasa.
- Hughes, A. (2003). *Testing for language teachers 2nd ed*. Cambridge: Cambridge University Press.
- Hyland, Ken dan Fiona Hyland. (2006). *Feedback in second language writing: contexts and issues*. Cambridge University Press.
- Jacobs, H, et al. (1981). *Testing ESL composition: A practical approach*. Rowley, MA: Newburry House.
- Kroll, Barbara. (1998). "Assessing writing abilities". Annual review of applied linguistics, Volume 18. Foundations of Second Language Teaching. USA: Cambridge University Press.
- Long, M. (1996). "The role of the linguistics environment in second language acquisition". *Handbook of second language acquisition*. W. Ritchie dan T. Bathia (ed.). San Diego: Academic Press.
- Lovorn, Michael G. dan Ali Reza Rezaei. (2011). "Assessing the assessment: Rubrics training for pre-service and new in-service teachers". *Practical Assessment, Research, and Evaluation*.
- Luoma, Sari. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.
- Mertler, Craig A. (2001). "Designing scoring rubrics for your classroom". *Practical Assessment, Research & Evaluation*, 7(25).

- McNamara, T.F. (1996). *Measuring second language performance*. London: Longman.
- McNamara, T.F. (2002). *Language testing*. Oxford: Oxford University Press.
- Moskal, B.M. (2000). "Scoring rubrics: What, when, and how". *Practical Assessment, Research & Evaluation*, 7(3).
- North, Brian. (2000). "Defining a flexible common measurement scale: Descriptors for self and teacher assessment". *Learner-directed assessment in ESL*. Glayol Ekbatani dan Herbert Pierson (ed.). New Jersey: Lawrence Erlbaum Associates.
- Nunan, D. (1992). *Research methods in language learning*. Cambridge: Cambridge University Press.
- Nunan, David dan Kathleen M. Bailey. (2009). *Exploring second language classroom research: A comprehensive guide*. USA: Heinle.
- Underhill, N. (1987). *Testing spoken language*. Cambridge: Cambridge University Press.
- Raimes, Ann. (1998). "Teaching writing". Annual Review Of Applied Linguistics. Volume 18, 1998. Foundations Of Second Language Teaching. USA: Cambridge University Press.
- Reddy, Malini Y. (2010). "Design and development of rubrics to improve assessment outcomes". ProQuest.
- Sammeng, Andi Mappi. (1996). "Pengajaran bahasa Indonesia sebagai bahasa asing serta peranannya." *Kumpulan sambutan dan makalah kongres internasional pengajaran bahasa Indonesia untuk penutur asing (BIPA)*. Depok: Fakultas Sastra Universitas Indonesia.
- Shaw, Stuart D. dan Cyril J. Weir. (2007). *Examining writing: Research and practice in assessing second language writing*. UK: Cambridge University Press.
- Silva, Tony dan Paul Kei Matsuda. (2001). *On second language writing*. New Jersey: Lawrence Erlbaum Associates.
- Stevens, Dannele dan Antonia J. Levi. (2005). *Introduction to rubrics: An assessment tool to save grading time, convey effective feedback, and promote student learning*. Virginia: Stylus Publishing.

- Taufik, Alvin. (2008). Penyusunan skema penilaian untuk kelas talking english tingkat dasar di ILP. Tesis yang belum diterbitkan.
- Weigle, S.C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Weir, Cyril J. (2005). *Language testing and validation: An evidence-based approach*. Great Britain: CPI Antony Rowe, Chippenham and Eastbourne.



Lampiran 1: Borang yang Diujikan

Nama siswa:

SKOR	BOBOT	PENGUNAAN BAHASA
	22-25	Penggunaan kalimat sederhana dan kompleks yang efektif. Mahir dalam pembentukan kata menggunakan afiksasi. Sedikit kesalahan dalam penggunaan konjungsi, kata ganti, dan preposisi.
	17-21	Sedikit kesalahan dalam penggunaan kalimat sederhana dan kompleks. Kadang terjadi kesalahan afiksasi. Beberapa kesalahan dalam penggunaan konjungsi, kata ganti, dan preposisi. Makna masih jelas.
	12-16	Banyak kesalahan dalam penggunaan kalimat sederhana atau kompleks. Banyak terjadi kesalahan pemakaian afiksasi. Banyak kesalahan dalam penggunaan konjungsi, kata ganti, dan preposisi. Makna membingungkan dan tidak jelas.
	1-11	Tidak menguasai pembentukan kalimat. Sangat banyak kesalahan dalam penggunaan afiksasi, konjungsi, kata ganti, dan preposisi. Tidak bermakna.
SKOR	BOBOT	GAGASAN DAN KREATIVITAS
	12-15	Mampu mengembangkan gagasan melebihi standar. Mampu memberikan contoh-contoh, ilustrasi, data, atau fakta untuk mendukung gagasan utama.
	8-11	Mengembangkan gagasan sesuai standar yang diminta. Sedikit memberikan contoh, ilustrasi, data, atau fakta untuk mendukung gagasan utama.
	4-7	Kurang bisa mengembangkan gagasan sesuai standar. Memberikan contoh, ilustrasi, data, dan fakta, tapi tidak sesuai untuk mendukung gagasan utama.
	1-3	Sangat kurang mengembangkan gagasan. Tidak memberikan contoh, ilustrasi, data, atau fakta untuk mendukung gagasan.
SKOR	BOBOT	ORGANISASI
	18-20	Tulisan terorganisasi (ada pembuka, isi, dan penutup). Penyampaian ide secara singkat, jelas, dan runtut.
	14-17	Tulisan kurang terorganisasi, tapi pikiran utama masih dapat dimengerti. Ide disampaikan secara jelas, tapi kurang ringkas dan tidak runtut.
	10-13	Tulisan kurang terorganisasi, tapi pikiran utama tidak dapat dimengerti karena penyampaian ide yang terputus-putus dan tidak lengkap.
	1-9	Tulisan sama sekali tidak terorganisasi dan tidak runtut.
SKOR	BOBOT	KOSAKATA
	12-15	Kosakata luas. Pemakaian kata yang efektif. Penggunaan ragam (formal, informal, ragam tulis, dsb) sangat tepat.
	8-11	Kosakata cukup luas. Kadang terjadi kesalahan pemakaian kata, tapi tidak mengurangi makna tulisan. Penggunaan ragam cukup tepat.
	4-7	Kosakata terbatas. Banyak terjadi kesalahan pemakaian kata sehingga makna membingungkan dan tidak jelas. Penggunaan ragam tidak konsisten.
	1-3	Kosakata terbatas. Pemakaian kata banyak yang salah. Penggunaan ragam tidak tepat.
SKOR	BOBOT	CAPAIAN TUGAS
	12-15	Tulisan relevan dengan topik yang diminta. Tulisan sesuai dengan instruksi tugas. Wawasan pengetahuan mengenai topik yang diberikan luas.
	8-11	Tulisan mayoritas relevan dengan topik, tapi kurang detail. Tulisan mayoritas sesuai dengan instruksi tugas. Wawasan pengetahuan mengenai topik cukup luas.
	4-7	Tulisan kurang relevan dengan topik dan tidak sepenuhnya sesuai dengan instruksi. Wawasan mengenai topik terbatas.
	1-3	Tulisan tidak relevan sama sekali dengan topik dan instruksi. Tidak memiliki pengetahuan tentang topik.
SKOR	BOBOT	EJAN DAN TANDA BACA
	8-10	Penguasaan yang baik. Sedikit terjadi kesalahan dalam ejaan, tanda baca, dan huruf kapital.
	4-7	Terkadang salah dalam ejaan, tanda baca, dan huruf kapital.
	1-3	Banyak salah dalam ejaan, tanda baca, dan huruf kapital.

Total skor:

Komentar:

Lampiran 2: Borang yang Telah Diperbaiki Hasil Wawancara

Nama siswa:

SKOR	BOBOT	PENGGUNAAN BAHASA
	22-25	Penggunaan kalimat sederhana dan kompleks yang efektif. Mahir dalam pembentukan kata menggunakan afiksasi. Sedikit kesalahan dalam penggunaan konjungsi, kata ganti, dan preposisi.
	17-21	Sedikit kesalahan dalam penggunaan kalimat sederhana dan kompleks. Kadang terjadi kesalahan afiksasi. Beberapa kesalahan dalam penggunaan konjungsi, kata ganti, dan preposisi. Makna masih jelas.
	12-16	Banyak kesalahan dalam penggunaan kalimat sederhana atau kompleks. Banyak terjadi kesalahan pemakaian afiksasi. Banyak kesalahan dalam penggunaan konjungsi, kata ganti, dan preposisi. Makna membingungkan dan tidak jelas.
	1-11	Tidak menguasai pembentukan kalimat. Sangat banyak kesalahan dalam penggunaan afiksasi, konjungsi, kata ganti, dan preposisi. Tidak bermakna.
SKOR	BOBOT	GAGASAN DAN KREATIVITAS
	12-15	Mampu mengembangkan gagasan melebihi standar. Mampu memberikan contoh-contoh, ilustrasi, data, atau fakta untuk mendukung gagasan utama.
	8-11	Mengembangkan gagasan sesuai standar yang diminta. Sedikit memberikan contoh, ilustrasi, data, atau fakta untuk mendukung gagasan utama.
	4-7	Kurang bisa mengembangkan gagasan sesuai standar. Memberikan contoh, ilustrasi, data, dan fakta, tapi tidak sesuai untuk mendukung gagasan utama.
	1-3	Sangat kurang mengembangkan gagasan. Tidak memberikan contoh, ilustrasi, data, atau fakta untuk mendukung gagasan.
SKOR	BOBOT	ORGANISASI
	18-20	Tulisan terorganisasi (ada pembuka, isi, dan penutup). Penyampaian ide secara singkat, jelas, dan runtut.
	14-17	Tulisan kurang terorganisasi, tapi pikiran utama masih dapat dimengerti. Ide disampaikan secara jelas, tapi kurang ringkas dan tidak runtut.
	10-13	Tulisan kurang terorganisasi, tapi pikiran utama tidak dapat dimengerti karena penyampaian ide yang terputus-putus dan tidak lengkap.
	1-9	Tulisan sama sekali tidak terorganisasi dan tidak runtut.
SKOR	BOBOT	KOSAKATA
	12-15	Kosakata luas. Pemakaian kata yang efektif. Penggunaan ragam (formal, informal, ragam tulis, dsb) sangat tepat.
	8-11	Kosakata cukup luas. Kadang terjadi kesalahan pemakaian kata, tapi tidak mengurangi makna tulisan. Penggunaan ragam cukup tepat.
	4-7	Kosakata terbatas. Banyak terjadi kesalahan pemakaian kata sehingga makna membingungkan dan tidak jelas. Penggunaan ragam tidak konsisten.
	1-3	Kosakata terbatas. Pemakaian kata banyak yang salah. Penggunaan ragam tidak tepat.
SKOR	BOBOT	CAPAIAN TUGAS
	12-15	Tulisan relevan dengan topik yang diminta. Tulisan sesuai dengan instruksi tugas. Wawasan pengetahuan mengenai topik yang diberikan luas.
	8-11	Tulisan mayoritas relevan dengan topik, tapi kurang detail. Tulisan mayoritas sesuai dengan instruksi tugas. Wawasan pengetahuan mengenai topik cukup luas.
	4-7	Tulisan kurang relevan dengan topik dan tidak sepenuhnya sesuai dengan instruksi. Wawasan mengenai topik terbatas.
	1-3	Tulisan tidak relevan sama sekali dengan topik dan instruksi. Tidak memiliki pengetahuan tentang topik.
SKOR	BOBOT	EJAJAN DAN TANDA BACA
	8-10	Penguasaan yang baik. Sedikit terjadi kesalahan dalam ejaan, tanda baca, dan huruf kapital.
	4-7	Terkadang salah dalam ejaan, tanda baca, dan huruf kapital.
	1-3	Banyak salah dalam ejaan, tanda baca, dan huruf kapital.

Total skor:

Catatan
Sedikit: < 40%
Sedang/kadang: 40-70%
Banyak: > 70%

Komentar:

Lampiran 3: Transkrip Wawancara Partisipan 1

Catatan transkrip:

1. Beberapa bagian transkrip penulis edit agar lebih formal, seperti kata *gw* penulis ganti dengan kata *saya*, kata *gimana* penulis ganti dengan *bagaimana*, dan sebagainya.
2. P adalah Penulis, sedangkan P1 adalah Partisipan satu, P2 adalah Partisipan 2, dan P3 adalah Partisipan 3.

P: Kenapa bisa paling rendah untuk UTS?

P1: Pertama ejaan. Kurang banyak kalimatnya dibandingkan yang lain yang sudah memberi banyak hasil. Saya cuma melihat ejaan dia di sini saja, tidak berkembang. Saya mau melihat bagaimana dia mengembangkan ejaannya. Ini berhubungan sama gagasannya.

Saya menilainya sedikit karena saya melihat data yang sedikit, karena itu berpengaruh bagaimana cara saya menilai dia.

P: Bagaimana dengan penentuan *range* nilai pada borang ini? Memakai *feeling* atau kira-kira?

P1: Kira-kira.

P: Bagaimana dengan penggunaan kata-kata seperti *sedikit* dan *banyak* di borang ini?

P1: Menurut saya *sedikit* itu 50 persen ke bawah, *banyak* 50 persen ke atas. Tergantung persepsi orang sedikit-banyaknya.

P: Tapi malah menimbulkan kebingungan atau tidak?

P1: Tidak, *sih*, malah mempermudah. Sebenarnya '*kan* intinya sama, tapi di sini ada *sedikit*, *lebih sedikit*, di sini *lebih banyak*. Tapi yang suka bikin bingung di sini, *sering* atau *kadang*. Apakah ini sering atau tidak ya. Sering itu berapa kali? Kadang-kadang berapa kali? Tergantung juga dari tulisan yang dihasilkan.

(lanjutan)

P: Kalau penggunaan kata *luas*, *cukup luas*, dan *terbatas*?

P1: Untuk itu '*kan* sudah dijelaskan dengan deskripsinya. Menurut saya, ini (deskripsi level) itu penjelasan dari kosakata cukup luas ini.

P: Anda jarang memberi nilai maksimal ya untuk tiap level?

P1: Mungkin begini. Misalnya dari satu sampai tiga saya beri nilai dua. Mungkin dia sering salah, tapi tidak terlalu sering. Jadi, saya beri dua.

P: Jadi menurut Anda rentang nilai itu membantu atau tidak?

P1: Membantu, terutama untuk rentang nilainya.

P: Dari boranganya sendiri apa kelebihan dan kekurangannya?

P1: Kelebihannya, ini '*kan* dibagi, *nih*, ya. Maksudnya dari enam aspek ini dijelaskan lagi ya, dari nilai segini sampai segini, ada levelnya. Jadi, ada gambaran.

Membantu lebih cepat juga. Tapi kadang-kadang kita harus benar-benar baca dulu, memahami dulu. Kalau tidak paham sekali, ya lama juga karena kita harus baca dulu. Jadi harus benar-benar dipahami dulu, baru, *deh*, bisa.

P: Kalau dibandingkan sama punya BIPA bukannya lebih cepat menggunakan punya BIPA ya?

P1: Ya makanya kalau kita sudah mengerti sekali ini, hafal di luar kepala, itu mungkin bisa lebih mudah dan cepat. Sebenarnya lebih mudah ya, kalau cepat, *kan*, tergantung dosennya.

P: Kalau pembagian ini, ada enam kriteria, terlalu banyak tidak?

P1: Tidak ya, hampir sama.

(lanjutan)

P: Menurut Anda kekurangan borang ini apa?

P1: Ini, deskripsinya panjang sekali. Awal-awal saya bingung. Ini saya harus benar-benar baca karena perbedaannya juga tipis, '*kan*, antara *sedikit*, *banyak*, *tidak*, juga kata *kadang* dan *sering*.

Mungkin kalau sudah terbiasa akan lebih mudah. Sebenarnya aspek-aspeknya itu penting ya.

P: Kalau deskripsinya ini sendiri sudah pas belum sama kriterianya? Apa ada yang redundan?

P1: Tidak, *sih*, tidak ada masalah untuk deskripsi tiap levelnya.

P: Kalau bobotnya? Rentang nilainya, apakah terlalu sedikit atau banyak?

P1: Tidak, *sih*, saya tidak melihat itu. Jadi saya menyesuaikan antara tulisan.

Justru lebih enak tipis daripada banyak. Kalau banyak itu lebih susah menentukannya. *Range*-nya jauh.

P: Menurut Anda hubungan borang ini dengan tulisan esai dan paragraf bagaimana?

P1: '*Kan* ini *range*-nya tinggi, levelnya banyak. Jadi lebih enak kalau mengoreksi tulisan yang banyak, seperti esai. Kalau yang pendek itu, levelnya itu terlalu banyak, jadi susah.

P: Ada saran atau masukan untuk borang ini?

P1: Ini, terlalu banyak katanya. Juga yang ini, kata-kata *sering* dan *kadang*.

Kalau ini berguna *deh*, yang deskripsi *memberi contoh*, *sedikit memberi contoh*. Berguna sekali.

Kurang cocok untuk paragraf. Karena begini, levelnya itu banyak, jadi kalau paragrafnya sedikit, ketahuan sekali kurangnya. Jadi itu alasan kenapa diberi nilai kecil.

Lampiran 4: Transkrip Wawancara Partisipan 2

P: Ada hambatan dari pemakaian borang yang kemarin tidak?

P2: Tidak ada masalah, *sih*, memakai borang ini, lancar-lancar saja. Mungkin agak bingung di kategori ejaan karena hanya ada tiga level, sedangkan yang lain empat level. Tidak apa-apa, *sih*, hanya kadang-kadang memberi nilainya bingung saja harus masuk yang mana.

Baik-baik saja, *kok*.

P: ada perbedaan cara pakai borang tidak untuk menilai UTS dan UAS?

P2: Ada, *kok*, di organisasi pasti. Menilainya lebih susah daripada menilai tulisan yang lebih panjang.

P: Borang ini mudah tidak dipakainya?

P2: Gampang, *kok*, benar. Memang awal-awal harus bolak-balik membaca deskripsi borangnya, tapi paling satu sampai tiga kali, setelah itu biasa.

Lebih mudah memakai borang ini daripada pakai yang di BIPA. Kalau di sini sudah ada level-levelnya, jadi mudah untuk menggolongkannya.

(Untuk hasil transkrip wawancara yang lebih lengkap dapat menghubungi penulis)

Lampiran 5: Contoh Borang yang telah Digunakan

Nama siswa: *Yutoro Sanada / 28*UTS/~~UAS~~

SKOR	BOBOT	PENGUNAAN BAHASA
16	22-25	Penggunaan kalimat sederhana dan kompleks yang efektif. Mahir dalam pembentukan kata menggunakan afiksasi. Sedikit kesalahan dalam penggunaan konjungsi, kata ganti, dan preposisi.
	17-21	Sedikit kesalahan dalam penggunaan kalimat sederhana dan kompleks. Kadang terjadi kesalahan afiksasi. Beberapa kesalahan dalam penggunaan konjungsi, kata ganti, dan preposisi. Makna masih jelas.
	12-16	Banyak kesalahan dalam penggunaan kalimat sederhana atau kompleks. Sering terjadi kesalahan pemakaian afiksasi. Sering salah dalam penggunaan konjungsi, kata ganti, dan preposisi. Makna membingungkan dan tidak jelas.
	1-11	Tidak menguasai pembentukan kalimat. Sangat banyak kesalahan dalam penggunaan afiksasi, konjungsi, kata ganti, dan preposisi. Tidak bermakna.
SKOR	BOBOT	GAGASAN DAN KREATIVITAS
8	12-15	Mampu mengembangkan gagasan melebihi standar. Mampu memberikan contoh-contoh, ilustrasi, data, atau fakta untuk mendukung gagasan utama.
	8-11	Mengembangkan gagasan sesuai standar yang diminta. Sedikit memberikan contoh, ilustrasi, data, atau fakta untuk mendukung gagasan utama.
	4-7	Kurang bisa mengembangkan gagasan sesuai standar. Memberikan contoh, ilustrasi, data, dan fakta, tapi tidak sesuai untuk mendukung gagasan utama.
	1-3	Sangat kurang mengembangkan gagasan. Tidak memberikan contoh, ilustrasi, data, atau fakta untuk mendukung gagasan.
SKOR	BOBOT	ORGANISASI
18	18-20	Tulisan terorganisasi (ada pembuka, isi, dan penutup). Penyampaian ide secara singkat, jelas, dan runtut.
	14-17	Tulisan kurang terorganisasi, tapi pikiran utama masih dapat dimengerti. Ide disampaikan secara jelas, tapi kurang ringkas dan tidak runtut.
	10-13	Tulisan kurang terorganisasi, tapi pikiran utama tidak dapat dimengerti karena penyampaian ide yang terputus-putus dan tidak lengkap.
	1-9	Tulisan sama sekali tidak terorganisasi dan tidak runtut.
SKOR	BOBOT	KOSAKATA
8	12-15	Kosakata luas. Pemakaian kata yang efektif. Penggunaan ragam (formal, informal, ragam tulis, dsb) sangat tepat.
	8-11	Kosakata cukup luas. Kadang terjadi kesalahan pemakaian kata, tapi tidak mengurangi makna tulisan. Penggunaan ragam cukup tepat.
	4-7	Kosakata terbatas. Sering terjadi kesalahan pemakaian kata sehingga makna membingungkan dan tidak jelas. Penggunaan ragam tidak konsisten.
	1-3	Kosakata terbatas. Pemakaian kata banyak yang salah. Penggunaan ragam tidak tepat.
SKOR	BOBOT	CAPAIAN TUGAS
8	12-15	Tulisan masih relevan dengan topik yang diminta. Tulisan sesuai dengan instruksi tugas. Wawasan pengetahuan mengenai topik yang diberikan luas.
	8-11	Tulisan mayoritas relevan dengan topik, tapi kurang detail. Tulisan mayoritas sesuai dengan instruksi tugas. Wawasan pengetahuan mengenai topik cukup luas.
	4-7	Tulisan kurang relevan dengan topik dan tidak sepenuhnya sesuai dengan instruksi. Wawasan mengenai topik terbatas.
	1-3	Tulisan tidak relevan sama sekali dengan topik dan instruksi. Tidak memiliki pengetahuan tentang topik.
SKOR	BOBOT	EJAJAN DAN TANDA BACA
7	8-10	Penguasaan yang baik. Sedikit terjadi kesalahan dalam ejaan, tanda baca, dan huruf kapital.
	4-7	Terkadang salah dalam ejaan, tanda baca, dan huruf kapital.
	1-3	Sering salah dalam ejaan, tanda baca, dan huruf kapital.

Total skor: *65*

Lampiran 6: Contoh Tulisan UTS

4

V. Pilihlah satu dari tiga pilihan di bawah ini! Kemudian, buatlah sebuah paragraf dari pilihan tersebut! Lingkari pilihan Anda! (Bobot: 50)

- a. Topik : KDRT
Metode pengembangan : definisi luas
- b. Topik : cara membuat mi instan
Metode pengembangan : proses
- c. Topik : hal-hal yang menyebabkan banjir di Jakarta
Metode pengembangan : sebab-akibat/kausalitas

~~Ka~~ Baru-baru ini, banyak orang di Indonesia makan mi instan. Makanan tersebut tidak hanya makanan itu yang enak tetapi juga makanan itu yang gampang dimasak. Cara memasak ada di bawah ini.


Pertama-tama, menyiapkan panci, air bersih, dan mi instan. Kemudian, dulu memanaskan air itu, ketika air itu direbus memasukkan mi instan itu ke dalam air.

Biasanya 5 menit kemudian, kami bisa makan mi instan yang sudah matang. Bukankah ini gampang!

Seperti itu, karena cara memasak mi instan gampang, ~~banyak~~ banyak orang suka makan mi instan.

Lampiran 7: Contoh Tulisan UAS

9/5/2011



UJIAN AKHIR SEMESTER MENULIS BIPA 2

SENIN, 9 MEI 2011 NAMA: Yutaro Sanada (2-B)

Bacalah situasi berikut.

Anda sedang berada di ruang tunggu Bandara Soekarno-Hatta. Di sana, Anda berkenalan dengan B. Anda bertanya tujuan perjalanan B. Ternyata, tujuan perjalanan B sama dengan tujuan perjalanan Anda. Hotei tempat B menginap juga sama dengan Anda. B bertanya kepada Anda tempat-tempat wisata yang menarik untuk dikunjungi. Anda menawarkan diri untuk menemani B berjalan-jalan di sana karena tujuan Anda ke kota itu juga untuk berwisata.

A. B

Buatlah sebuah tulisan yang di dalamnya terdapat deskripsi diri/orang dan deskripsi kegiatan berdasarkan situasi di atas. Berilah judul yang menarik untuk tulisan tersebut. Ketika menulis gunakan ejaan, tanda baca, kalimat, dan kepaduan paragraf yang baik. Buat minimal 5 paragraf.

Pertemuan Kebetulan

① Nama saya Yutaro Sanada. Saya pegawai swasta sekaligus mahasiswa BIPA UI. Mengapa saya belajar bahasa Indonesia di UI? Alasannya, perusahaan saya yang bernama Mitsubishi mengirintak saya ke Indonesia karena negara Indonesia adalah negara yang penting bagi Mitsubishi. Alasannya, Indonesia kaya akan sumber alam, misalnya migas. Biasanya,

(lanjutan)

Saya sibuk karena harus belajar di UI sampai jam 12.30, lalu harus pergi ke kantor. Namun, saya dapat mengambil cuti ketika Lebaran.

③ Saya memutuskan berkunjung ke Kalimantan Timur ketika Lebaran. Ketika saya menunggu pesawat di ruang tunggu Bandara Soekarno-Hatta, saya berkenalan dengan perempuan yang sangat cantik seperti bidadari. Namanya Maria.

Saya terkejut karena perempuan yang berkuliah fakultas ekonomi UI itu juga akan berwisata ke Kalimantan Timur bahkan penghitungannya juga sama dengan saya! Saya bersyukur kepada Tuhan. Saya harus mempertahankan diri dan mempertahankan kesihatan yang dapat dilakukan di sana kepada Maria untuk menemani Maria berjalan-jalan di sana.

④ Hotel tempat^o yang kita inap bernama "Alice" ada di Samarinda. Setelah menikmati makanan khas di hotel itu, kita dapat mengunjungi wilayah konsentrasi untuk orang utan yang bernama "Sambaja". Dari Alice ke Sambaja memakan waktu kurang lebih (2) jam dengan menggunakan bus. Di Sambaja, kita dapat melihat orang utan dari jarak dekat, tetapi tidak boleh memesang untuk mencegah penularan penyakit, seperti

(lanjutan)

hepatitis dari manusia ke orang utan. Seperti yang kita ketahui, kita harus memelihara orang utan karena orang hutan adalah hewan langka. Walaupun tidak boleh membunuh orang utan yang tua itu, kita cukup menikmati suasana Samboja.

④ Setelah menikmati Samboja, kita bertolak ke sungai Mahkam. Sungai Mahkam ada di barat dari Samarinda kira-kira 3 jam dengan bus. Di sana ada suku dayak. Suku dayak adalah suku yang berbeda dengan suku yang lain. Suku yang tinggalnya panjang itu membela kehidupan yang unik. Namun, jumlah suku yang gaya hidupnya unik itu terus-menerus turun. Alasannya, urbanisasi di sana sedang maju.



(lanjutan)

Banyak perusahaan datang ke sana untuk mendapatkan kayu yang kualitasnya bagus. Oleh karena itu, banyak suku Dayak terutama pemuda mengikuti kegiatan yang tren supaya dapat mencari nafkah. Walaupun demikian, kita masih dapat bertemu dengan suku Dayak kalau berkunjung ke sana.

⑤ Seperti yang saya jelaskan tersebut, kita dapat melakukan kegiatan yang menarik di Kalimantan Timur. Kita dapat melihat hewan langka yang tinggal di hanya Kalimantan dan Sumatra, bahkan dapat bertemu dengan suku yang unik. Namun, sekiranya saja sedikit keredaan karena dapat berenang, berjemur, berselancar dan lain-lain kalau kita berwisata ke Pulau Bali. Mudah-mudahan saya berwisata ke Pulau Bali bersama Miffa nanti setelah saya berpacaran silaturahmi dengan Miffa.