

PENERAPAN ANALYTICAL SCORING DAN HOLISTIC SCORING PADA MATA KULIAH PENULISAN DESKRIPTIF BAHASA MANDARIN

Livia Vasantadjaja

Fakultas Bahasa dan Budaya, Program Sastra China, Universitas Kristen Maranatha

Jl. Prof. drg. Surya Sumantri, M.P.H. No. 65

Bandung - 40164, Jawa Barat, Indonesia

Telp: +62 22 - 201 2186

E-mail: livia.vasantadjaja@let.maranatha.edu

Abstrak

Dalam pendidikan Bahasa Mandarin, kemampuan menulis menjadi salah satu kemampuan berbahasa yang penting dan harus dikuasai dengan baik. Kemampuan menulis memang memiliki tingkat kesulitannya sendiri, misalnya dalam Bahasa Mandarin, pengetahuan dan kemampuan menulis aksara Han menjadi kemampuan dasar yang harus dimiliki oleh pemelajar Bahasa Mandarin, kemudian baru diikuti dengan kemampuan merangkai kata, kalimat, paragraf menjadi sebuah wacana yang utuh. Oleh sebab itu, materi drilling atau tes menulis menjadi penting, serta memiliki instrumen penilaian dan penilai yang reliabel sangat diperlukan untuk mencapai keberhasilan pembelajaran. Untuk itu, penelitian ini akan menerapkan dan menguji reliabilitas dua jenis instrumen penilaian holistic scoring dan analytical scoring yang digunakan untuk menilai setiap karangan deskriptif bahasa Mandarin yang dihasilkan oleh mahasiswa prodi Sastra China Universitas Kristen Maranatha tingkat II sepanjang satu semester. Melalui usaha dua orang penilai dalam menerapkan dua jenis instrumen penilaian ini, diharapkan dapat menghasilkan instrumen penilaian yang konsisten dan dapat dijadikan bahan evaluasi pembelajaran yang tepat dan akurat.

Kata Kunci: *holistic scoring, analytical scoring, bahasa Mandarin, menulis, reliabilitas*

Abstract

In Chinese language, writing skills are one of important language skills that should be mastered. Writing skills undeniably have their own difficulties, for example, in Chinese language, learners should have the knowledge and ability to write Han characters before they are able to compose Chinese writings, then followed by other skills, such as vocabulary, grammar, organization skills of writing. Therefore, drilling materials or writing tests, rating scales, as well as reliable raters become essential. This study will examine the reliability of two rating scales, which are holistic scoring dan analytical scoring, assessed by two raters in Chinese Descriptive Writing course in Chinese Department of Maranatha Christian University. The purpose of this study is to show the result of writing assessment using these two writing scales, to show reliability and consistency between two raters in assessing writing. So that, this study will help and improve the assessment quality of Chinese writings courses to be more objective, reliable, consistent and accurate.

Keywords: *holistic scoring, analytical scoring, Chinese language, Chinese writing, reliability*

A. PENDAHULUAN

Dalam pendidikan dan pengajaran Bahasa Mandarin, menulis (写作) merupakan kemampuan berbahasa yang lebih kompleks dan sulit dibandingkan dengan menyimak, membaca, dan berbicara. (Yang, 2010, p. 238; Zhao, 2005, p.140; Zhou, 2009, p.433) Menurut Zhou Xiaobin & Liu Hai'ou (2004), kesulitan menulis dalam Bahasa Mandarin disebabkan oleh tiga alasan berikut:

- a. Menulis termasuk dalam sistem penuturan tertulis yang perlu dipelajari.
- b. Menulis dalam bahasa Mandarin menggunakan aksara Han yang merupakan aksara piktoğraf, bukan menggunakan huruf alfabet.
- c. Menulis sebagai alat komunikasi secara tertulis, akan direspons atau dinyatakan kembali melalui tulisan. Berbeda dengan berbicara yang bisa direspons secara langsung dan bisa menggunakan cara lain seperti gestur tangan, gerakan tubuh, dan lainnya.

Oleh sebab itu, dalam menulis diperlukan tingkat kemampuan berbahasa yang lebih tinggi dibandingkan dengan kemampuan yang lain. Kemampuan menulis membutuhkan perpaduan pengetahuan aksara Han, kosakata, tata bahasa, dan wacana. Selain itu, kemampuan menulis juga perlu didukung kemampuan berpikir, kemampuan observasi, latar belakang dan pengetahuan budaya, minat, dan sebagainya. Bagi penutur asli yang tidak memiliki kesulitan dalam berbahasapun, kompleksitas dalam hal menulis membuat kemampuan untuk menyatakan hal secara tertulis dengan baik ini hanya dikuasai sedikit orang. (Zhou, 2009, p. 434)

Untuk melatih kemampuan menulis dalam Bahasa Mandarin maka pemelajar memerlukan latihan untuk menulis. Tujuan utama dari latihan menulis ini adalah untuk mempraktekkan berbagai cara penuturan secara tertulis, diperlukan perbaikan secara kontinu terhadap kesalahan-kesalahan yang terjadi ketika menulis, dengan demikian akan semakin meningkatkan kemampuan menulis pemelajar. Contoh latihan menulis yang sering diterapkan dalam pembelajaran Bahasa Mandarin: 听后写 menuliskan kembali apa yang didengar, 看后写 menuliskan kembali apa yang diilustrasikan melalui gambar, foto, peta, komik, tabel data, dan lain-lain, 读后写 menuliskan kembali apa yang dibaca, 续写 meneruskan cerita.

Kemampuan menulis pemelajar Bahasa Mandarin merefleksikan level kemampuan berbahasa pemelajar itu sendiri, yang mencakup ketepatan penggunaan tata bahasa, kosakata, tanda baca, menyusun paragraf menjadi sebuah wacana. Tulisan atau karangan itu memerlukan *feedback* atau evaluasi untuk melihat sejauh mana kemampuan menulis pemelajar Bahasa Mandarin. Untuk itu seorang pengajar atau penilai memerlukan waktu yang lebih panjang untuk membaca dan memeriksa karangan, dan kemungkinan adanya subjektivitas dari penilai yang sulit dihindari. Dua hal ini menjadi kelemahan atau masalah dalam penilaian untuk menulis, sehingga diperlukan adanya metode atau cara yang objektif baik dari segi penyajian materi atau dari segi penilaian, di antaranya (Liu, 2007, p.392):

- a. Standar penilaian yang lebih detail dan terukur. Standar penilaian mencakup isi karangan, struktur wacana, kesalahan penulisan aksara, penggunaan tanda baca, ketentuan deduksi nilai dan unsur lainnya yang terkait. Semakin detail standar yang ditetapkan, maka akan semakin objektif hasil penilaiannya. Para penilai harus mengikuti standar penilaian yang telah ditetapkan.

- b. Karangan yang sama dinilai oleh beberapa orang penilai. (多人集体平分)
- c. Setiap penilai memberikan penilaian hanya pada salah satu unsur standar di setiap tulisan. Nilai akhir yang diperoleh adalah gabungan nilai dari beberapa penilai yang menilai unsur linguistik yang berbeda. (流水评分)
- d. Penilaian dilakukan secara menyeluruh pada karangan dan penilai cukup mendapatkan kesan yang menonjol dari setiap karangan, kemudian dibandingkan dengan karangan yang lain. (比较评分)
- e. Penyajian materi menulis menyediakan ketentuan yang spesifik, seperti instruksi untuk menulis jenis karangan yang harus disajikan, panjang karangan dan lainnya.

Faktor subjektivitas dan durasi penilaian yang memakan waktu bisa diatasi dengan ketersediaan instrumen penilaian dan penilai yang objektif, penyajian materi menulis yang spesifik. Untuk mengatasi hal tersebut yang juga terjadi dan dialami dalam proses belajar dan mengajar di dalam perkuliahan menulis, maka penulis ingin membahas perangkat penilaian yang digunakan dalam mata kuliah Penulisan Teks Deskriptif Bahasa China tingkat II di program studi Sastra China Universitas Kristen Maranatha. Pada level ini, mahasiswa dianggap telah menguasai Bahasa Mandarin tingkat dasar dan akan memulai pembelajaran Bahasa Mandarin tingkat menengah. Pemilihan ini juga dikarenakan kesempatan yang diberikan kepada penulis untuk menjalankan pengajaran dan pendidikan di level ini sebagai bagian dari Tridharma Perguruan Tinggi.

Menulis adalah salah satu dari empat kemampuan berbahasa yang dinilai sulit karena menuntut pemelajar untuk mengaplikasikan semua kemampuan berbahasa lain. Dalam Bahasa Mandarin, kemampuan menulis ini harus didukung oleh kemampuan menulis aksara Han yang benar, memilih diksi yang tepat, merangkai frasa atau kalimat dengan struktur tata bahasa yang tepat, cara penyampaian sebuah kalimat, pemakaian gaya bahasa tertentu, menjadi tantangan di dalam mengarang dalam Bahasa Mandarin. Dengan banyaknya komponen kemampuan yang harus dilatih, *drilling* dalam mata kuliah Penulisan Teks Deskriptif menjadi kunci keberhasilan dalam mempelajari aspek berbahasa ini. Banyak latihan menulis yang harus dilakukan, dimulai dari merangkai kalimat, menyusun kalimat menjadi paragraf, menulis sebuah wacana, dan lain-lain. Hasil dari setiap *drilling* berupa karangan ini tentu memerlukan penilaian yang objektif sebagai bahan evaluasi pembelajaran.

Untuk itu, penulis menggunakan dua instrumen penilaian, yakni *holistic scoring* dan *analytical scoring* untuk menilai karangan dalam Bahasa Mandarin. Dua instrumen penilaian ini adalah perangkat penilaian yang sering digunakan dalam penilaian menulis/mengarang untuk mengukur kemampuan berbahasa asing secara tertulis dalam bentuk esai/karangan. Penggunaan dan perbandingan dua instrumen penilaian ini akan mengurangi subjektivitas dalam hal pemberian nilai pada sebuah tulisan, dan untuk mengukur keakuratan hasil pencapaian pembelajaran.

Selain itu, untuk mengatasi subjektivitas, penulis juga menyediakan dua orang penilai yang akan menilai setiap karangan dengan menggunakan *holistic scoring* dan *analytical scoring* yang telah disediakan. Kemudian hasil penilaian dari setiap karangan yang dihasilkan mahasiswa sepanjang satu semester dari kedua penilai dengan menggunakan dua instrumen penilaian akan ditampilkan dan dibandingkan. Akan terlihat bagaimana konsistensi hasil dari penilaian yang dilakukan oleh para penilai

dengan dua instrumen *holistic scoring* dan *analytical scoring* melalui uji reliabilitas yang akan dilakukan.

Dengan perbandingan dua jenis instrumen penilaian yang digunakan bertujuan untuk menelaah hasil dari *holistic scoring* dan *analytical scoring*, dapat dilihat seberapa reliabel hasil penilaian yang dihasilkan dari dua penilai tersebut. Karena faktor reliabilitas menjadi hal yang paling utama dalam mengembangkan perangkat tes berbahasa, seperti soal tes, instrumen penilaian dan lainnya. Instrumen penilaian ini telah diterapkan sebelumnya oleh seorang penilai dalam menilai karangan, akan tetapi tidak diketahui jelas evaluasi dari penerapan instrumen ini, unsur reliabel, validitas dari instrumen ini tidak bisa terukur. Maka dari itu, menilai mata kuliah mengarang dalam Bahasa Mandarin perlu langkah lanjutan, yaitu menerapkan dua jenis instrumen penilaian dengan penilai lebih dari satu orang, dan mencermati seberapa reliabel penerapan dua jenis instrumen penilaian oleh dua orang penilai ini. Semakin reliabel semakin instrumen penilaian ini dapat digunakan oleh lebih banyak penilai dan dapat dipakai dengan cakupan yang lebih luas.

Diharapkan telaah ilmiah tahap awal ini dapat menjadi acuan pembaharuan dan evaluasi secara kontinu dalam melakukan pembaharuan instrumen penilaian, merancang dan membuat instrumen penilaian yang terukur sesuai dengan tujuan pembelajaran, dan dapat mendeskripsikan kemampuan pemelajar Bahasa Mandarin, terutama untuk mata kuliah menulis atau mengarang dalam Bahasa Mandarin.

B. TINJAUAN PUSTAKA

2.1 Tes Bahasa

Dalam siklus penerapan model pembelajaran bahasa asing, tes bahasa menjadi satu bagian penting untuk mengukur kemampuan berbahasa pemelajar dan menilai hasil pembelajaran kemampuan berbahasa. Sheng Yan mendefinisikan tes sebagai cara untuk mengukur kemampuan dan pengetahuan seseorang dalam suatu bidang tertentu. Di dalam sebuah tes, terdapat beberapa kriteria yang harus dipenuhi, di antaranya: (1) tes harus ada wujud atau tampilan atau kegiatan atau aktivitas; (2) tes harus memiliki tujuan pengukuran yang jelas; (3) tes harus dapat mengukur kemampuan dan pengetahuan seseorang; (4) tes harus dapat mengungkapkan kemampuan dan pengetahuan seseorang walaupun hanya melalui sampel dari tes tersebut; (5) tes dapat mengukur ruang lingkup atau aspek yang spesifik. (Sheng, 1989, pp 333-334) .

Banyak ragam tes yang sering diterapkan dalam pengajaran dan pembelajaran bahasa asing. Ditinjau dari sisi kegunaan atau fungsi tes, maka jenis tes dapat dikategorikan menjadi lima jenis, yaitu: *achievement test* 成绩测试, *aptitude test* 学能测试, *diagnostic test* 诊断性测试, *placement test* 分班测试, dan *proficiency test* 能力测试. (Sheng, 1989, pp 335-336) Dari berbagai jenis tes tersebut, *achievement test* merupakan jenis tes yang paling banyak digunakan.

Achievement test bertujuan mengukur penguasaan materi dan kemampuan pemelajar dalam materi pembelajaran tertentu. Keunggulan *achievement test* adalah materi yang telah dipelajari atau diajarkan akan menjadi materi yang diuji; materi tes ditentukan oleh pengajar; memiliki sifat reflektif sehingga bisa dimanfaatkan untuk acuan evaluasi pembelajaran untuk pengajar dan pemelajar. (Sheng, 1989, p.335; Yang, 2010, p. 2) Bentuk tes ini yang sering dijumpai berupa ujian tengah dan akhir semester.

Dalam *achievement test* terdapat jenis tes yang memiliki lingkup yang lebih kecil, yaitu *progress test* atau *progress achievement test* 进步测试. Jenis tes ini dilakukan secara berkala dan konsisten, dilengkapi dengan rancangan pembelajaran dan penyediaan materi yang menunjang. Tes ini digunakan untuk mengukur kemampuan penguasaan pemelajar pada setiap materi atau bab yang dipelajari. Contoh tes ini bisa berupa latihan, kuis atau tes kecil. Dengan demikian pengukuran pencapaian pembelajaran pada setiap pemelajar dan kemajuan dari setiap pemelajar pada setiap bab materi/unit dapat terukur lebih akurat.

2.2 Tes Menulis dalam Bahasa Mandarin 汉语写作测试

Tes menulis Bahasa Mandarin ialah tes untuk mengetahui dan mengukur kemampuan pemelajar dalam mengungkapkan atau mengutarakan hal secara tertulis dalam Bahasa Mandarin. Kemampuan menulis dibandingkan dengan kemampuan menyimak, membaca dan berbicara memiliki tingkat kesulitan yang paling tinggi, karena menulis membutuhkan kemampuan berbahasa yang lain, seperti menulis aksara Han; penggunaan kosakata, tata bahasa; menulis dan mengkonstruksi paragraf, wacana; pemilihan gaya penulisan dan gaya bahasa secara baik dan benar. (Yang, 2010, p. 8)

Yang Yi (2010) pun menyatakan beragam bentuk materi untuk menguji kemampuan menulis dalam Bahasa Mandarin, contohnya seperti 听后写 menuliskan kembali hal yang didengar, 看图写 menulis sesuai dengan gambar, 看影视后写 menuliskan kembali hal yang dipertunjukkan/divisualisasikan, 扩写 mengembangkan karangan, 缩写 meringkas karangan, 改写 memperbaiki karangan, 续写 melanjutkan karangan; atau bentuk karangan yang berdasarkan fungsi atau kegunaannya, seperti menulis memo, surat, jurnal harian, dokumen resmi; atau bentuk karangan berdasarkan isinya, seperti menulis tentang figur/tokoh/orang, peristiwa, pemandangan, lokasi, benda/binatang; atau berdasarkan jenis karangan, seperti naratif, deskriptif, argumentatif dan lainnya. Bentuk-bentuk karangan tersebut tidak hanya diperuntukkan sebagai tes atau materi yang diujikan, tetapi juga bisa digunakan sebagai materi latihan menulis.

Dalam prinsip penilaian tes menulis, ada beberapa komponen yang harus diperhatikan, yaitu:

- Ketepatan menulis aksara Han;
- Ketepatan dalam penggunaan kosakata, pembentukan kalimat sesuai dengan kaidah tata bahasa yang baik dan benar;
- Kesenambungan dan keterkaitan antarkalimat dan paragraf;
- Kejelasan makna dan inti yang diutarakan secara tertulis dalam Bahasa Mandarin;
- Variasi kalimat yang ditampilkan;

- Ketepatan dan kesesuaian format penulisan dan penggunaan tanda baca.

Beberapa hal tersebut dalam proses pembelajaran menulis harus dilatih dan dikuasai, karena melibatkan pengetahuan dan kemampuan berbahasa yang kompleks dan menyeluruh, seperti kemampuan menulis aksara Han, kemampuan merangkai kata-kata, penguasaan tata bahasa Mandarin, kemampuan menyusun kalimat menjadi paragraf dan sebagainya.

2.3 Holistic Scoring dan Analytic Scoring

Dalam menilai kualitas tulisan atau karangan dalam pembelajaran bahasa asing, ketersediaan perangkat penilaian yang tepat sangat diperlukan, terlebih lagi untuk menghindari penilaian subjektif dari penilai. Untuk membangun penilaian yang objektif dan reliabel dapat diterapkan beberapa hal, misalnya: kejelasan instruksi materi uji menulis, penetapan durasi untuk menulis, tema karangan, panjang karangan, struktur penulisan, jenis karangan yang harus dibuat, instrumen penilaian, dan lainnya.

Ada dua jenis instrumen penilaian yang lazim digunakan dalam menilai kemampuan menulis. Yang pertama, *holistic scoring* atau *impressionistic scoring*. *Holistic scoring* adalah penilaian yang didasarkan pada kesan keseluruhan. Penilaian jenis ini banyak digunakan untuk penilaian dengan massa besar. (Hughes, 2005, pp 94-95)

Kelebihan dari penilaian dengan *holistic scoring* ini adalah:

- a. Kecepatan dan singkatnya waktu dalam memberikan penilaian.
- b. Fokus penilaian langsung pada bagian atau aspek yang paling menonjol dari tulisan tersebut, sehingga penilai langsung melihat kelebihan yang dinyatakan dalam tulisan.

Di samping itu, *holistic scoring* juga memiliki kekurangan, yaitu:

- a. Skor yang diberikan oleh penilai tidak memberikan diagnosa yang rinci tentang kemampuan menulis seseorang.
- b. Skor yang diberikan tidak mudah diinterpretasikan. Kesamaan skor mungkin menggambarkan penekanan penilaian aspek menulis yang berbeda.

Jenis kedua, *analytical scoring* adalah tipe *scoring* yang mempertimbangkan banyak aspek di dalamnya, seperti tata bahasa, kosakata, diksi, koherensi dalam karangan, dan lain-lain. (Hughes, 2005, p. 100) Maka dari itu, *analytical scoring* membutuhkan konsep yang jelas dalam proses perancangan perangkat penilaian, dan membutuhkan waktu yang lebih banyak untuk mempertimbangkan detail-detail komponen penilaian yang perlu diperhatikan.

Analytical scoring memiliki kelebihan, di antaranya:

- a. Penilaian akan cenderung meniadakan hal-hal yang tidak seimbang dalam perkembangan kemampuan berbahasa seseorang.
- b. Penilai harus mempertimbangkan beberapa aspek yang sebenarnya ingin dihindari atau diabaikan ketika penilaian berlangsung.
- c. Penilai harus memberikan poin untuk setiap aspek dan menjadikan penilaian yang diberikan dapat dipercaya.

Selain kelebihan di atas, *analytical scoring* pun memiliki kekurangan, di antaranya:

- a. Durasi waktu yang diperlukan untuk menilai lebih panjang dibandingkan dengan *holistic scoring*.

- b. Fokus penilai pada beberapa aspek penilaian membuat perhatian penilai teralihkan dari keutuhan dan kesatuan sebuah tulisan.

Pemilihan dan penggunaan instrumen penilaian dari dua jenis *scoring* ini tergantung dari tujuan diselenggarakannya sebuah tes, jumlah pemelajar, jumlah penilai, jenis karangan yang harus ditulis dan lainnya. Dengan adanya dua jenis penilaian ini, aspek-aspek untuk dijadikan pertimbangan dalam memberikan penilaian dapat menjadi evaluasi yang tepat bagi sebuah pembelajaran bahasa.

2.4 Reliabilitas

Reliabilitas menjadi komponen pertama yang disebutkan oleh Bachman & Palmer ketika mengutarakan konsep "*test of usefulness*" sebagai pertimbangan penting dalam rancangan dan pengembangan tes bahasa, selain komponen yang lainnya, yakni: *construct validity, authenticity, interactiveness, impact, dan practicality*. (dalam Weigle, p.48)

Reliabilitas didefinisikan sebagai konsistensi hasil dalam pengukuran atau penilaian dalam berbagai situasi atau kondisi. (Weigle, 2009, p.49; Zhang, 2005, p.99; Sheng, 1989, p.444) Sheng Yan (1989) memaparkan bahwa reliabilitas, pengukuran yang bersifat reliabel berarti ketika mengukur atau menilai sesuatu menggunakan standar yang sama dan jikalau dilakukan berulang kali dalam berbagai situasi dan kondisi, maka nilai yang dihasilkan tetap sama. Dalam konteks penilaian berbahasa, maka hal ini menunjukkan hasil tes yang konsisten dan stabil. Contoh, ketika memberikan penilaian terhadap sebuah karangan, penilai pertama memberikan nilai 90, penilai kedua memberikan nilai 90, maka dapat disimpulkan bahwa ada realibilitas dalam penilaian; kalau penilai pertama memberikan nilai 90, penilai kedua memberikan nilai 70, maka hal ini menunjukkan tidak ada reliabilitas di dalamnya. (pp 444-445)

Reliabilitas dalam mengukur atau menguji kemampuan menulis tentunya akan dipengaruhi oleh beberapa unsur, seperti unsur yang berkait dengan tugas/instruksi dalam menulis (topik, respons dalam menulis); unsur yang berkait dengan kemampuan berbahasa pemelajar; unsur yang berkait dengan proses penilaian/*scoring* (latar belakang, pengalaman penilai, skala penilaian/pengukuran yang digunakan, pelatihan bagi penilai).

Dalam proses penilaian kemampuan menulis, adanya instrumen penilaian seperti *analytical scoring* dan *holistic scoring* akan sangat membantu dalam menghasilkan hasil yang cukup objektif dan reliabel. Ada beberapa cara untuk mengukur reliabilitas atau konsistensi antarpemilai yang menggunakan *analytical scoring* dan *holistic scoring* untuk menilai tulisan. Dua aspek penting dalam reliabilitas menurut Weigle (2009) adalah:

- a. *Intra-rater reliability (self-consistency)*: kecenderungan seorang penilai memberikan skor yang sama pada teks yang sama dalam kondisi yang berbeda.
- b. *Inter-rater reliability (agreement between raters)*: kecenderungan penilai yang berbeda memberikan skor yang sama pada teks yang sama.

Dengan mempertimbangkan dua kondisi aspek di atas, reliabilitas dapat dihitung dengan menggunakan koefisien korelasi. Hasil perhitungan koefisien korelasi akan menunjukkan angka antara 0 dan 1, yang menggambarkan seberapa kuat hubungan antara skor yang dinilai. Jika mendekati nol, maka semakin kecil atau tidak adanya

relasi di antara penilai, sedangkan jika angka mendekati satu, maka menunjukkan seberapa kuat relasi antarpemilai. Untuk penghitungan korelasi bisa digunakan dengan metode Spearman atau Pearson.

2.5 Penelitian terkait

Pembahasan tentang aspek penilaian kemampuan berbahasa seperti yang diterapkan dalam tulisan ini banyak sekali ditemukan dalam penelitian dengan lingkup pengajaran dan pendidikan Bahasa Inggris, salah satunya seperti penelitian perbandingan *holistic scoring* dan *analytical scoring* yang dilakukan oleh Yuji Nakamura dari Tokyo Keizai University dengan judul *A Comparison of Holistic and Analytic Scoring Methods in the Assessment of Writing* yang dimuat dalam *The Interface Between Interlanguage, Pragmatics and Assessment: Proceedings of the 3rd Annual JALT Pan-SIG Conference* tahun 2004. Dalam makalahnya, Nakamura menelaah perbandingan kelebihan dan kekurangan dari dua instrumen penilaian *holistic scoring* dan *analytical scoring* melalui secara kuantitatif dengan berdasarkan pada konsep enam komponen "test of usefulness", yaitu *reliability, construct validity, authenticity, interactiveness, impact, practicality* dari Bachman & Palmer yang diadaptasi oleh Wieggle. Dengan adanya perbandingan dua instrumen penilaian ini maka akan diketahui instrumen mana yang lebih baik untuk menilai karangan pemelajar Bahasa Inggris.

Untuk konteks dan ruang lingkup pendidikan dan pengajaran Bahasa Mandarin, beberapa penelitian menyangkut perangkat penilaian berbahasa untuk menulis atau mengarang Bahasa Mandarin mungkin tidak sebanyak dalam Bahasa Inggris, seperti ditulis oleh Ma Mingxia yang berjudul *对外汉语三种写作测验方式的效度实证研究* (2012) (*Duiwai Hanyu San Zhong Xiezu Ceyan Fangshi de Xiaodu Shizheng Yanjiu*), penelitian yang membuktikan kesahihan dari tiga jenis ujian menulis yang muncul dalam HSK; selain itu penelitian lain yang berjudul *Testing Writing in Chinese as a Second Language: An Overview of Research* (2016) yang ditulis oleh Zhang Dongbo, Li Li dan Zhao Shouhui yang berfokus pada penilaian kemampuan mengarang yang ada pada HSK.

Penelitian lain yang serupa membahas tentang menulis/mengarang Bahasa Mandarin lainnya cenderung mengambil objek penelitian tentang perangkat penilaian yang dilakukan pada jenis tes berupa *proficiency tests*, bukan pada *achievement tests*, terlihat dalam beberapa artikel ilmiah dalam *Chinese as a Second Language Assessment* (2017), seperti: *Exploring the Relationship Between Raters' Personality Traits and Rating Severity in Writing Assessment of Chinese as a Second Language: A Pilot Study*; dan *Computer-Mediated Corrective Feedback in Chinese as a Second Language Writing: Learners' Perspectives*.

C. METODE PENELITIAN

Penelitian ini menerapkan metode penelitian deskriptif dengan pendekatan kuantitatif. Metode deskriptif adalah metode penelitian yang ditujukan untuk mengungkapkan fenomena-fenomena yang aktual, penggambaran fenomena-fenomena bisa secara individu maupun kelompok, dan menggunakan angka-angka. Dengan pendekatan kuantitatif, fenomena-fenomena objektif dikaji secara kuantitatif dengan menggunakan statistik, struktur dan percobaan terkontrol. (Hamdi &

Bahrudin, 2014, pp 5-6) Dengan metode penelitian yang sistematis, terstruktur dan terukur, maka kematangan konsep sangat krusial.

Untuk pelaksanaan penelitian ini, rencana pembelajaran semester untuk mata kuliah penulisan teks deskriptif menjadi titik awal merancang konsep dan mempersiapkan instrumen dan data sepanjang satu semester. Sampel penelitian adalah 10 set karangan yang merupakan *drilling* dan dikategorikan sebagai *progress assessment tests* dari 11 mahasiswa tingkat II, sehingga seluruh sampel yang digunakan dalam penelitian ini berjumlah 110 karangan berupa wacana dengan panjang setiap karangan 200-250 aksara Han.

Berikut ini adalah materi latihan (*drilling*) yang diberikan kepada mahasiswa:

No	Materi Latihan
1	<p>听录音后写作:</p> <p>一天晚上有一个人慌慌张张地来到警察局,对警察说:“你们,你们能帮我找到我妻子吗?我从早上9点就没有看见她了,你们必须帮我找到她!”</p> <p>“您妻子长得什么样?”警察问。</p> <p>“不知道。”</p> <p>“她有多高?”</p> <p>“不太清楚。”</p> <p>“那她穿的什么衣服?”</p> <p>“我忘了,我只知道她走的时候带了一条狗。”</p> <p>“狗?什么样的狗?”</p> <p>“白色的,身上有五块黑色的花,有一只脚是黑的,高63公分,体重21公斤,早上出去以前还没吃东西呢。”</p> <p>“好,您放心,我们一定帮您找到那条狗。”</p>
2	<p>读后写作:</p> <p>我是一名记者,身上总是带着很多笔,因为一支笔很快就用完了。有一天,我的笔有都用完了,正好同事小李要去商店,我就请他帮我买笔,我对他说:“请你帮我买12支圆珠笔,我不喜欢黑色的,不要黑色,你一定要记住,别买黑色的。”</p> <p>过了一会儿,他回来了,也给我买来了笔,我打开一看,十二支全是黑色的!我对他说:“我不是说不要黑色的吗?”没想到小李说:“你说了好几遍黑色、黑色、黑色、黑色这个词印象太深了,我就记住了黑色,所以一进商店我就找黑色的。”</p> <p>哎,这也不能怪他。要是我请他买笔的时候,说话简单一点儿:“请你帮我买12支圆珠笔,蓝色的。”我想这样的话,他大概就不会买错了。</p> <p>从那以后说话或者写文章,我都是能简单就简单,不说没有的话。</p>
3	<p>续写:</p> <p>(结尾) 我爱你……</p>
4	<p>扩写:</p> <p>(开头) 一般来说,我周末都没有什么重要的事儿。……………</p>

5	人物描写： 人生中最有影响或最重要的人。
6	人物性格描写： 你梦中的男友或女友。
7	动物描写： 
8	事物描写：（看视频后写作） 
9	处所描写：我的家乡。
10	情景描写：去看音乐会。

Selain menyediakan materi penulisan karangan deskriptif Bahasa Mandarin, juga perlu mendesain dua perangkat instrumen penilaian berdasarkan *holistic scoring* dan *analytical scoring (4-rating scale)*, seperti berikut:

分数	项目
4	文章的主要内容、思想很明确、通顺的。 文章条理、组织清楚。 用词正确。 有极少数语法错误。 无或极少汉字错误。
3	文章的主要内容、思想很明确、通顺的。 文章组织适当清楚。 使用词语良好。 比较少有语法错误。 有少数汉字错误。
2	文章的主要内容、思想已表明，但不清楚。 文章没有条理，组织不那么好。 词汇使用一般。 有一些主要语法错误。 有一些汉字错误。
1	文章的主要内容、思想很难识别。 文章条理、组织得不好。 使用词语很有限。 许多语法错误。 许多汉字错误。
0	无内容 无思路 文章组织混乱 几乎所有词语、语法使用不当 几乎所有的汉字不准确

Tabel 3.1: Instrumen Penilaian *Holistic Scoring*

	4	3	2	1	0
内容	内容描述出有趣、有一致性、清楚的思想。	内容描述相当清楚、有一致性的思想。	内容描述有些不清楚的思想。	内容思想、思路不清晰。	- 无内容, 无思路 - 文章组织混乱 - 几乎所有词语、语法使用不当 - 几乎所有的汉字不准确
组织	内容组织十分清楚, 连贯性很强, 话题描述深刻。	内容组织合乎逻辑, 关于主题的大部分信息包括在内。	有些地方不合逻辑, 松散的组织, 令人困惑, 缺乏主题信息。	组织断开, 很差。主题不完全覆盖。	
词汇	使用丰富的词汇, 其使用正确, 尤其是描述性的词语。 (有 1-2 错误)	使用各种相关的词汇。使用的词汇符合其主题。 (有 3-4 错误)	使用的一些词语是基本的和重复的。 (有 5-6 错误)	没有足够的词汇, 词汇低于预期水平, 通常与主题无关。 (有 ≥ 7 错误)	
语法	使用正确的语法、词序、标点符号。在复句上有极少错误。 (有 1-2 错误)	有比较少错误的语法、词序、标点符号。 (有 3-4 错误)	使用语法、词序、标点符号有一些错误, 影响到理解。 (有 5-6 错误)	许多语法、词序、标点符号的错误, 尤其基本结构的错误。 (有 ≥ 7 错误)	
汉字	具有 1-2 汉字错误	具有 3-4 汉字错误	具有 5-6 汉字错误	具有 ≥ 7 汉字错误	

Tabel 3.2: Instrumen Penilaian *Analytical Scoring*

Perancangan dua jenis instrumen penilaian berdasarkan komponen kemampuan berbahasa yang dinyatakan oleh Bachman & Palmer (dalam Weigle, 2009, pp42-43), yaitu komponen pengetahuan bahasa, pengetahuan tekstual, pengetahuan fungsional, dan pengetahuan sociolinguistik. Dalam konteks perancangan instrumen penilaian untuk menulis atau mengarang dalam Bahasa Mandarin, penulis menitikberatkan pada komponen pengetahuan bahasa dan tekstual, di dalamnya mencakup pengetahuan dalam penulisan aksara Han, penggunaan kosakata, tata bahasa, dan pengorganisasian isi karangan.

Penilaian terhadap sampel 110 karangan dengan dua jenis instrumen penilaian ini dilaksanakan oleh dua orang penilai (*raters*) yang *non-native speakers* yang memiliki pengalaman dalam pendidikan dan pengajaran Bahasa Mandarin, memiliki gelar magister, memiliki kemampuan berbahasa setara dengan *New HSK level 6*. Penilaian dilakukan masing-masing penilai di tempat yang berbeda. Dalam penyajian dan analisis data akan menampilkan seluruh hasil penilaian, perhitungan koefisien korelasi Pearson untuk melihat konsistensi penilaian antarpemilai terhadap seluruh karangan yang ada.

D. HASIL DAN PEMBAHASAN

Arthur Hughes mengatakan bahwa cara terbaik untuk menguji kemampuan menulis seseorang adalah dengan membuatnya menulis (2005, p. 83) Hal inilah yang dilakukan dalam mata kuliah Penulisan Teks Deskriptif Bahasa Mandarin tingkat II, bukan hanya untuk menguji kemampuan menulis, tapi juga untuk berlatih menyampaikan hal-hal secara tertulis. Mata kuliah ini bertujuan agar mahasiswa memiliki kemampuan menyampaikan hal-hal yang ditangkap melalui panca indera, sehingga pembaca dapat melihat, mengalami, merasakan hal yang sama dengan menggunakan Bahasa Mandarin secara tertulis. Misalnya, mahasiswa mendeskripsikan salah satu anggota keluarganya, atau mendeskripsikan tentang kampung halaman, atau mendeskripsikan sebuah acara/peristiwa.

Dalam proses kegiatan menulis dalam pembelajaran bahasa asing, pengujian kemampuan menulis melibatkan dua komponen dasar, yakni satu atau lebih latihan/tes menulis, atau pemberian instruksi untuk menulis; dan alat untuk mengevaluasi hasil latihan/tes menulis. (Weigle, 2009, p.2) Maka dalam proses perancangan dan pengambilan data untuk penelitian ini dua komponen dasar ini selalu menjadi pertimbangan dasar.

Komponen pertama dalam pengujian kemampuan menulis Bahasa Mandarin di sini, terdiri dari *progress tests* atau *progress achievement tests* (进步测试) yaitu seluruh latihan/tes menulis deskriptif Bahasa Mandarin yang materi dan tujuan pembelajaran tiap pertemuannya telah ditentukan dalam rencana pembelajaran semester. Bentuk *drilling* atau tes menulis deskriptif Bahasa China berupa wacana ini juga seperti 听后写 menuliskan kembali hal yang didengar, 看图写 menulis sesuai gambar, 看影视后写 menuliskan kembali hal yang ditampilkan, 扩写 mengembangkan karangan, 续写 melanjutkan karangan dan lainnya. Tema penulisan deskriptif yang dilatih, seperti mendeskripsikan secara tertulis tentang anggota keluarga, benda, bangunan, lokasi, alam, dan lain-lain.

Materi-materi latihan disediakan untuk memberikan lebih banyak kesempatan untuk berlatih dalam menulis, serta didukung dengan evaluasi dari proses pembelajaran. Sehingga komponen dasar kedua yang dinyatakan oleh Weigle juga harus menjadi fokus terpenting. Untuk mengukur pencapaian tujuan pembelajaran, diperlukan instrumen penilaian yang terukur, detil dan objektif, yang memiliki komponen penilaian akan pengorganisasian karangan, isi karangan, penulisan aksara Han, pemilihan kosakata, penggunaan struktur tata bahasa, jenis format penulisan, penggunaan tanda baca.

Aspek-aspek penilaian di atas dituangkan dalam dua jenis rubrik penilaian yang lazim digunakan untuk menilai kemampuan menulis seseorang, yaitu *holistic scoring* dan *analytical scoring* (lihat tabel 3.1 & tabel 3.2). *Holistic scoring* yang menekankan pada penilaian yang bersifat global dan singkatnya waktu dalam mengoreksi; sedangkan *analytical scoring* menekankan pada penilaian yang lebih detail pada setiap komponen penilaian, dan akan menghasilkan evaluasi yang lebih spesifik, meskipun waktu yang diperlukan lebih banyak dibandingkan dengan *holistic scoring*.

Upaya menghasilkan penilaian yang objektif juga didukung dengan tersedianya dua orang penilai *non native speakers* yang memiliki kompetensi dalam bidang pengajaran Bahasa Mandarin. Dalam pembahasan ini, melalui dua orang penilai, dua jenis instrumen penilaian, dan 110 sampel karangan deskriptif Bahasa Mandarin yang dinilai masing-masing oleh penilai dengan instrumen *holistic scoring* dan *analytical scoring*, akan menguji reliabilitas (*inter-rater reliability*). Hasil penilaian tersebut akan dihitung dengan koefisien korelasi Pearson dengan rumus sebagai berikut:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

r = koefisien korelasi Pearson; x = hasil penilai pertama; y= hasil penilai kedua

Interval Koefisien	Tingkat Hubungan
+ .70 or higher	Very strong positive relationship
+ .40 to +.69	Strong positive relationship
+ .30 to +.39	Moderate positive relationship
+ .20 to +.29	weak positive relationship
+ .01 to +.19	No or negligible relationship
0	No relationship [zero correlation]
- .01 to -.19	No or negligible relationship
- .20 to -.29	weak negative relationship
- .30 to -.39	Moderate negative relationship
- .40 to -.69	Strong negative relationship
- .70 or higher	Very strong negative relationship

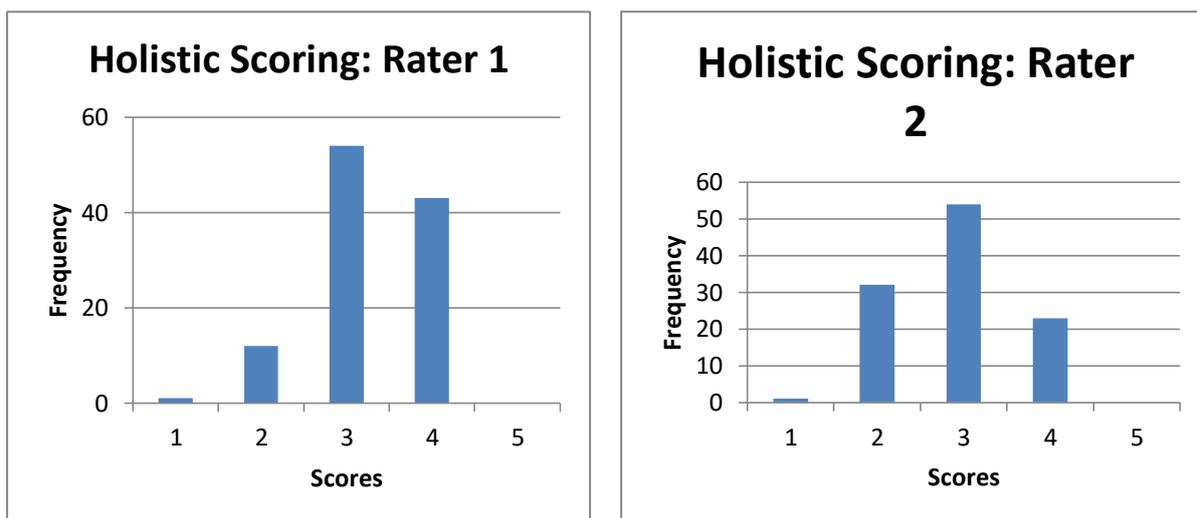
Tabel 4.1:
Pearson's Correlation Coefficients

Tabel 4.2:
Hasil Penilaian
Penulisan Teks
Deskriptif dengan
Holistic Scoring.

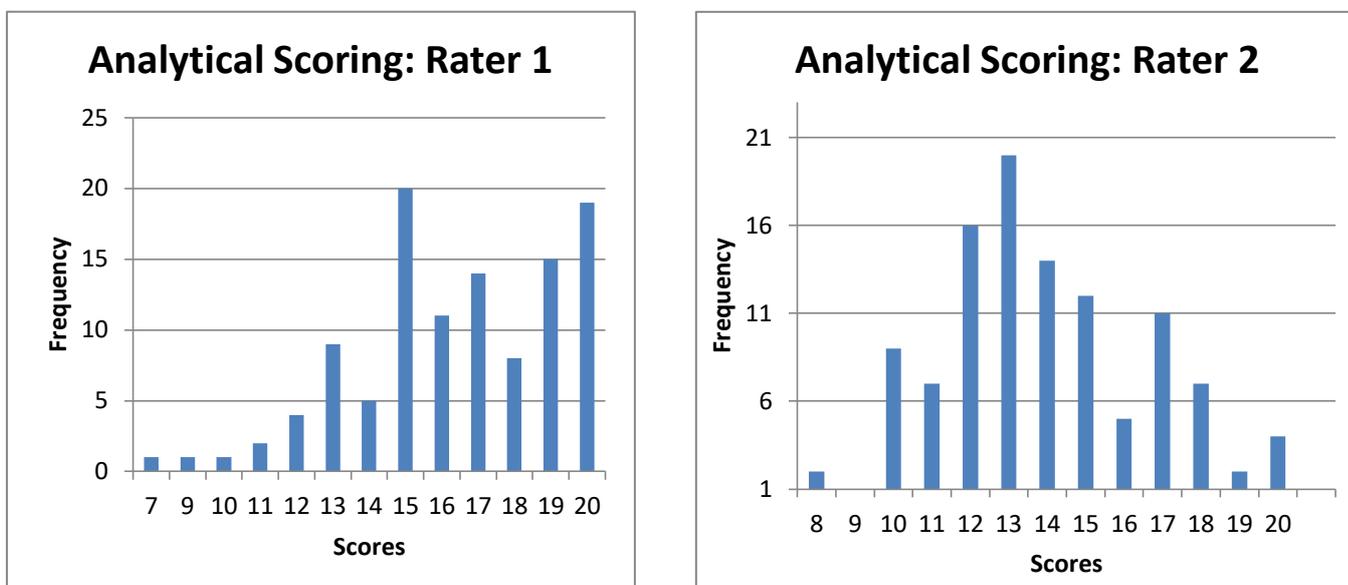
NO	HOLISTIC	
	RATER 1	RATER 2
1	3	2
2	4	4
3	2	2
4	4	3
5	4	2
6	2	3
7	3	4
8	4	3
9	2	3
10	4	4
11	3	3
12	2	2
13	3	2
14	2	2
15	2	3
16	3	2
17	2	1
18	3	3
19	3	2
20	3	3
21	4	4
22	3	2
23	3	3
24	4	4
25	3	3
26	3	4
27	3	3
28	3	2
29	3	3
30	4	4
31	3	2
32	4	4
33	3	3
34	3	2
35	4	3
36	3	2
37	4	3
38	3	2
39	3	2
40	3	3
41	2	4
42	2	2
43	3	4
44	4	4
45	3	2
46	4	4
47	4	2
48	4	3
49	4	3
50	3	3
51	4	3
52	3	3
53	3	2
54	4	4
55	4	4
56	4	3
57	4	4
58	3	3
59	2	4
60	3	3
61	1	2
62	4	3
63	4	4
64	2	3
65	3	4
66	4	3
67	4	2
68	4	4
69	3	3
70	3	3
71	3	2
72	4	3
73	3	3
74	4	3
75	3	3
76	4	4
77	3	3
78	3	3
79	4	4
80	3	2
81	4	3
82	3	2
83	4	3
84	3	3
85	3	3
86	3	2
87	4	3
88	4	3
89	3	2
90	4	3
91	3	2
92	3	3
93	3	2
94	3	2
95	4	3
96	3	2
97	3	2
98	4	3
99	4	3
100	2	3
101	4	4
102	3	3
103	3	3
104	3	3
105	3	3
106	4	3

Tabel 4.3:
Hasil Penilaian Penulisan
Teks Deskriptif dengan
Analytical Scoring.

NO	ANALYTICAL	
	RATER 1	RATER 2
1	16	10
2	18	17
3	13	10
4	17	14
5	20	13
6	10	13
7	17	18
8	20	14
9	12	14
10	17	18
11	15	13
12	13	10
13	17	12
14	13	9
15	12	13
16	14	8
17	9	8
18	13	15
19	17	11
20	15	11
21	17	18
22	17	12
23	17	16
24	19	20
25	16	15
26	16	17
27	13	13
28	17	12
29	16	15
30	20	17
31	15	13
32	20	20
33	16	17
34	16	10
35	19	15
36	15	10
37	19	15
38	15	12
39	14	11
40	15	15
41	13	20
42	11	12
43	16	17
44	20	18
45	18	12
46	19	19
47	20	12
48	20	14
49	20	15
50	15	13
51	19	16
52	15	14
53	14	12
54	19	17
55	20	16
56	19	13
57	20	19
58	15	13
59	12	17
60	15	13
61	7	12
62	19	14
63	19	17
64	11	14
65	17	18
66	18	14
67	20	10
68	19	20
69	15	13
70	15	15
71	16	12
72	19	13
73	17	13
74	18	13
75	13	16
76	20	17
77	15	14
78	17	12
79	20	18
80	13	10
81	20	13
82	14	10
83	20	14
84	16	15
85	16	13
86	15	12
87	20	15
88	19	15
89	15	12
90	20	17
91	15	11
92	18	17
93	15	10
94	17	12
95	18	14
96	16	11
97	15	11
98	20	14
99	19	13
100	12	13
101	20	18
102	15	13
103	13	15
104	14	12
105	15	14
106	19	14



Tabel 4.4: Histogram *Holistic Scoring* Penilai 1 dan Penilai 2



Tabel 4.5: Histogram *Analytic Scoring* Penilai 1 dan Penilai 2

Hasil Penilaian *Holistic Scoring*

		RATER 1	RATER 2
RATER 1	Pearson's Correlation	1.00	0.40
	Sig. (1-tailed)		.000
	N	110	110
RATER 2	Pearson's Correlation	0.40	1.00
	Sig. (1-tailed)	.000	
	N	110	110

Hasil Penilaian *Analytical Scoring*

		RATER 1	RATER 2
RATER 1	Pearson's Correlation	1.00	0.41
	Sig. (1-tailed)		.000
	N	110	110
RATER 2	Pearson's Correlation	0,41	1.00
	Sig. (1-tailed)	.000	
	N	110	110

Tabel 4.6: Hasil Uji Reliabilitas *Holistic Scoring* & *Analytical Scoring*

Tabel 4.2 sampai tabel 4.6 di atas merupakan penyajian kuantitatif hasil penilaian yang dilakukan oleh dua orang penilai terhadap 110 karangan dari mata kuliah Penulisan Teks Deskriptif Bahasa Mandarin dengan instrumen *holistic scoring* dan *analytical scoring*.

Hasil dari instrumen *holistic scoring* menghasilkan koefisien korelasi sebesar $r_1=0,40$ memiliki arti bahwa reliabilitas dari hasil penilaian tersebut kuat dan positif. Pada tabel 4.2, dapat diamati penilaian yang diberikan dua orang penilai memiliki *mean* atau rata-rata dari nilai yang didapat sebesar 3,26 dan 2,90. Skor yang paling sering muncul (*mode*) dan nilai tengah (*median*) pada instrumen jenis ini pada dua penilai sama, yakni skor 3. Standar deviasi dari penilaian satu dan dua adalah 0,15 dan 0,06.

Hasil dari instrumen *analytical scoring*, koefisien korelasi yang dimiliki $r_2 = 0,41$, yang berarti hubungannya kuat dan positif. Tabel 4.3, menunjukkan *mean* atau rata-rata nilai yang didapat dari penilai 1 dan penilai 2 adalah 16,42 dan 13,92, dengan standar deviasi sebesar 1,03 dan 0,56. Nilai skor yang paling sering muncul (*mode*) pada penilai 1 adalah skor 15, dan penilai 2 adalah skor 13. Nilai tengah (*median*) yang tampak pada hasil penilai 1 adalah skor 17 dan nilai tengah yang diberikan penilai 2 adalah skor 14.

Jika kita membandingkan kedua hasil instrumen penilaian dilihat dari koefisien korelasi Pearson $r_1 = 0,40$ dan $r_2 = 0,41$ maka perbandingan reliabilitas keduanya hampir sama. Secara umum kedua penilai memiliki hubungan konsistensi yang kuat dan positif baik dari penilai maupun instrumen yang dipakai untuk penilaian. Hasil pencapaian pemelajaran jika dilihat dari hasil penilaian *holistic scoring*, seluruhnya ada di skor 3, yang berarti ada di level baik; dari hasil penilaian *analytical scoring*, skor yang dicapai ada antara skor 14-17, yang berarti pencapaian pemelajaran ada di antara level cukup dan baik. Dengan hasil yang didapat dari dua instrumen penilaian ini reliabel dan konsisten, dan dapat digunakan untuk mengukur hasil capaian pemelajaran mahasiswa tingkat II dalam penulisan teks deskriptif Bahasa Mandarin.

E. KESIMPULAN

Setelah melakukan pengolahan, penghitungan dan analisis data secara kuantitatif berdasarkan koefisien korelasi Pearson, maka dapat disimpulkan penerapan *analytical scoring* dan *holistic scoring* sebagai instrumen penilaian pada mata kuliah Penulisan Teks Deskriptif Bahasa Mandarin tingkat II dinilai konsisten dan reliabel, ditunjukkan melalui koefisien korelasi masing-masing instrumen sebesar 0,40 dan 0,41. Melalui penghitungan ukuran tendensi sentral yang terdiri dari *mean*, *median*, *mode*, serta standar deviasi dapat dilihat pencapaian pemelajaran mahasiswa tingkat II pada mata kuliah Penulisan Teks Deskriptif Bahasa China telah tercapai dengan baik.

Penggunaan dua jenis instrumen penilaian ini ternyata tidak menyatakan perbedaan nilai uji realibilitas yang signifikan ($r_1 = 0,40$; $r_2 = 0,41$). Walaupun keduanya tidak mencapai level relasi sangat kuat dan positif ($r = 0,7-0,99$) yang menunjukkan semakin konsisten dan idealnya reliabilitas instrumen penilaian tersebut. Namun kedua instrumen ini menunjukkan hubungan yang kuat, positif dan konsisten, yang valid untuk digunakan. Dalam proses penilaian, kedua penilai pun menemukan kelebihan dan kekurangan dari dua instrumen penilaian ini. *Holistic scoring* sangat membantu dalam efisiensi waktu dan kecepatan menyelesaikan penilaian, namun penilaiannya pun berlaku secara umum. Sebaliknya, *analytical scoring* membuat penilai

membutuhkan durasi waktu lebih lama karena ada aspek kemampuan berbahasa yang harus dipertimbangkan dan tidak bisa diabaikan begitu saja, sehingga penilai bisa melihat kekuatan dan kelemahan aspek mana yang dimiliki setiap mahasiswa pada setiap materi karangan deskriptif. Hal ini sangat membantu evaluasi untuk peningkatan pembelajaran yang lebih terarah.

Dari uji reliabilitas penilaian yang dilakukan ternyata hasil akhir dari kedua penilai hampir sama, sehingga dalam studi kasus kali ini pemakaian instrumen *holistic scoring* atau *analytical scoring* tidak akan terlalu menimbulkan pengaruh signifikan dalam penilaian pencapaian pembelajaran. Namun seperti yang diungkapkan oleh Liu Xun (2007) dengan adanya perangkat standar penilaian yang detail, jumlah penilai lebih dari satu, materi yang dilengkapi dengan instruksi penulisan yang detail akan memunculkan penilaian kemampuan menulis yang objektif.

Setelah melakukan penelitian ini, penulis mendapatkan perlunya evaluasi yang lebih menyeluruh dan spesifik untuk pembelajaran menulis dalam Bahasa Mandarin di program studi Sastra China Universitas Kristen Maranatha, seperti pemutakhiran materi ajar, pembaharuan detail deskripsi skoring dalam instrumen penilaian, dan penyamaan konsep penilaian antarpenilai.

DAFTAR PUSTAKA

- [1] Bahruddin, E. & Hamdi, A.S., 2014. *Metode penelitian kuantitatif aplikasi dalam pendidikan*. Yogyakarta: Deepublish.
- [2] Dongbo, Zhang., Li, Li. & Shouhui, Zhao., 2016. *Testing Writing in Chinese as a Second Language: An Overview of Research* in Trends in Language Assessment Research and Practice: The View from the Middle East and the Pacific Rim, pp. 315-333, Cambridge: Cambridge Scholars Publishing.
- [3] Everson, M.E & Shen, Helen H. (2010). *Research Among Learners of Chinese as a Foreign Language* (Chinese Language Teachers Association Monograph Series, Vol. IV). Hawai'i: National Foreign Language Resource Center University of Hawai'i.
- [4] Everson, M.E. & Xiao, Yun, eds. (2009). *Teaching Chinese as a foreign language: theories and applications*. Boston: Cheng & Tsui Company, Inc.
- [5] Hughes, A. (2005). *Testing for language teachers*. Cambridge: Cambridge University Press.
- [6] Liu, Xun. (2007). *Duiwai hanyu jiaoyu xue yinlun*. Beijing: Beijing Language and Culture University Press.
- [7] Sheng, Yan. (1989). *Yuyan jiaoxue yuanli*. Beijing: Guojia Hanban.
- [8] Wallace, M.J. (2001). *Action research for language teachers*. Cambridge: Cambridge University Press.
- [9] Weigle, S. C. (2009). *Assessing writing*, Cambridge: Cambridge University Press.
- [10] Yang, Yi. (2010). *Duiwai hanyu jiaoxue de chengji cheshi*. Beijing: Peking University Press.
- [11] Yang, Huiyuan. (2007). *Ketang jiaoxue lilun yu shijian*. Beijing: Beijing Language and Culture University Press.

- [12] Zhang, Dongbo & Lin, Chin-Hsi. (2017). *Chinese as a second language assessment*. 10.1007/978-981-10-4089-4.
- [13] Zhang, Kai. (2005). *Yuyan ceyan lilun yu shijian*. Beijing: Beijing Language and Culture University Press.
- [14] Zhao, Jinming. (2006). *Hanyu keyi zheyang jiao—yuyan jineng pian*. Beijing: Shangwu Yinshu Guan.
- [15] Zhao, Jinming. (2005). *Duiwai hanyu jiaoxue gailun*. Beijing: Shangwu Yinshu Guan.
- [16] Zhou, Xiaobin & Li, Hai'ou. (2004). *Duiwai hanyu jiaoxue rumen*. Guangzhou: Zhongshan Daxue Chubanshe.
- [17] Zhou, Xiaobin. (2009). *Duiwai hanyu jiaoxue rumen (di'er ban)*. Guangzhou: Zhongshan Daxue Chubanshe.

DATA PENULIS

Nama : Livia Vasandtajaja
Program studi : Sastra China
Perguruan Tinggi : Fakultas Bahasa dan Budaya, Universitas Kristen Maranatha
Telp. : +62 22 201 2186
Email : livia.vasantadjaja@let.maranatha.edu