

BAB I

PENDAHULUAN

1.1 Latar Belakang

Menulis adalah proses kompleks yang penting dalam kaitannya dengan aktivitas belajar, berpikir, dan berkomunikasi dengan orang lain (Dunsmuir & Clifford, 2003). Menulis bukan hanya terkait dengan kemampuan berbahasa yang harus dikuasai, tetapi juga sudah menjadi kebutuhan yang harus dipenuhi oleh kaum intelektual dan akademisi, khususnya bagi mahasiswa, guru, dan peneliti. Bahkan, menulis karya tulis ilmiah dan memublikasikannya merupakan suatu kegiatan yang sifatnya wajib bagi mahasiswa sebagai salah satu syarat kelulusan mereka (Dirjen Dikti, 2012). Karena sangat penting dan menyangkut masa depan dan nasib seseorang, penilaian menulis, khususnya penilaian karya tulis ilmiah, harus bersifat adil dan objektif dan jangan sampai merugikan seseorang. Oleh karena itu, penilaian menulis harus dapat diandalkan, dapat dipercaya, dan menjadi prioritas utama yang harus dicapai dalam penilaian (American Psychological Association, 1999).

Salah satu unsur yang sangat penting dalam penilaian menulis, khususnya jenis karya tulis ilmiah (KTI) adalah instrumen penilaiannya. Pada instrumen penilaian menulis terdapat kriteria penilaian kualitas karya tulis yang secara substansial dapat meningkatkan keandalan penilaian karya tulis tersebut (Shohamy, 1992). Untuk memperoleh penilaian menulis yang adil dan objektif, diperlukan instrumen penilaian, yaitu rubrik KTI yang efektif. Rubrik yang efektif dipercaya dapat mengurangi subjektivitas, efek penilai, serta bias dalam suatu penilaian (Rezaei & Lovorn, 2010). Rubrik penilaian KTI dikatakan efektif jika memenuhi syarat-syarat tertentu, yaitu didesain dengan baik, bersifat kontekstual dan analitis, serta dilengkapi dengan panduan penjelasan (Rezaei & Lovorn, 2010). Selain itu, rubrik harus terperinci dan memenuhi desain tes yang baik, yaitu memenuhi validitas dan reliabilitas yang setiap item penilaiannya telah dikalibrasi dengan model tes termutakhir.

Kebutuhan akan instrumen penilaian karya tulis ilmiah yang memenuhi syarat validitas dan reliabilitas untuk berbagai keperluan akademis sangat tinggi.

Namun, berdasarkan pengamatan terbatas, masih sedikit rubrik penilaian KTI yang memenuhi syarat-syarat tersebut. Rubrik-rubrik penilaian KTI yang ada di Indonesia umumnya masih disusun berdasarkan konvensi setiap institusi atau pengelola jurnal ilmiah sehingga format dan kriterianya berbeda-beda antara satu dengan yang lain. Selain itu, tidak ada jaminan bahwa rubrik-rubrik penilaian itu telah divalidasi dengan baik dan layak sehingga tidak ada jaminan pula bahwa hasil penilaian artikel ilmiah yang dilakukan itu memenuhi prinsip keadilan dan objektivitas. Dengan terpenuhinya kriteria validitas dan reliabilitas rubrik, sangat dimungkinkan bahwa hal-hal yang terkait dengan subjektivitas penilai dalam penilaian artikel ilmiah akan teratasi dengan baik sehingga memaksimalkan objektivitas dalam penilaian menulis.

Untuk mencapai objektivitas dalam penilaian menulis, peran penilai juga sangat penting (Santos, 1988; Weigle, 1999). Dengan kata lain, keberhasilan penilaian menulis sangat ditentukan oleh penilai (Kane, Crooks, Cohen, 1996). Dalam penilaian menulis, konsistensi penilai sangat menentukan tercapainya reliabilitas penilaian. Konsistensi penilai mengacu pada tingkat kesepakatan antara penilai yang sama (konsistensi intrarater) atau penilai yang berbeda (konsistensi inter-rater) (Feinstein, 1987). Konsistensi penilai mengacu pada sejauh mana skor yang diberikan oleh penilai dengan nilai yang stabil dan bebas dari kesalahan (Yen, 2016).

Ketika melakukan penilaian menulis dalam tim, berkelompok, atau secara berpasangan, penilai harus konsisten agar hasil penilaian tersebut reliabel. Skor yang diberikan oleh penilai satu dengan penilai lainnya, untuk kinerja yang sama, hasilnya relatif sama, hampir sama, atau hanya memiliki perbedaan minimal (Weir, 2005). Masalahnya, penilaian menulis melibatkan penilai yang memiliki kekurangan, yaitu mengandung unsur subjektivitas (Messics, 1996). Subjektivitas penilai dalam penilaian menulis tersebut menimbulkan perbedaan skor antarpenilai (Schaefer, 2008), ketidakkonsistenan penilaian (Huang, 2008), atau variabilitas skor penilaian (Barkaoui, 2011).

Variabilitas antar-penilai adalah bagian yang tidak dapat dihindari dari proses penilaian menulis (Linacre, 1989). Oleh karena itu, hal yang dapat dilakukan pada variabilitas adalah dengan meminimalkan efeknya, bukan

menghilangkannya sama sekali. Untuk meminimalkan efek variabilitas tersebut, perlu diketahui wujud atau manifestasi dari variabilitas penilaian yang melibatkan penilai itu terlebih dahulu. Setelah diketahui wujudnya, akan dapat ditentukan metode apa yang harus dilakukan untuk meminimalkan efek variabilitas tersebut.

Variabilitas penilai dapat bermanifestasi atau terjadi dalam bentuk, yaitu (1) pengacuan pada rubrik penilaian, (2) penafsiran kriteria penilaian (3) pemahaman dan penggunaan skala penilaian, atau dalam (4) perbedaan tingkat keketatan atau kelonggaran penilaian (*severity-lenency*) (Weigle, 2002; Lumley, 2005). Dari beberapa perwujudan variabilitas penilaian tersebut, *severity* (untuk selanjutnya disebut *severitas*) merupakan manifestasi variabilitas yang paling dapat diidentifikasi dengan kentara karena wujudnya yang konkret, yaitu berupa skor atau angka yang dapat dilihat, dikenali, dan dihitung. Severitas penilai mengacu pada kecenderungan penilai untuk secara konsisten memberi nilai lebih tinggi atau lebih rendah daripada nilai yang seharusnya (Engelhard, 1994). Linacre (dalam Lumly dan McNamara, 1995) menggunakan istilah *severitas* untuk menyebut perbedaan antara penilai dalam cara mereka menafsirkan skala penilaian. Sebagai manifestasi dari variabilitas penilaian, *severitas* ini dapat dikatakan sebagai sumber kesalahan yang dapat memengaruhi kualitas penilaian menulis.

Ada beberapa faktor yang diduga dapat memengaruhi *severitas* penilai. Salah satunya adalah faktor kepribadian penilai (Bernadin dkk., 2000). Hal itu karena penilaian tidak terlepas dari pengaruh ide, preferensi, atau keyakinan dari si penilai itu sendiri (Alaei dkk., 2014). Sejumlah peneliti dan ahli teori berspekulasi bahwa kepribadian penilai berperan dalam proses penilaian kinerja. Landy dan Farr (1980) menyatakan bahwa perbedaan kepribadian penilai memengaruhi strategi penilaian kinerja penilai. Tziner dkk. (2005) juga mengemukakan bahwa ciri-ciri kepribadian memengaruhi perilaku penilai dalam proses penilaian. Lebih lanjut, sejumlah investigasi empiris telah dilakukan terkait dengan kepribadian penilai dalam penilaian kinerja (Randall & Sharples, 2012).

Namun, penelitian-penelitian itu menunjukkan hasil yang berbeda-beda sehingga sulit ditarik kesimpulan yang bermakna. Terdapat temuan yang tidak konsisten dalam beberapa literatur terkait permasalahan penilai ini. Misalnya,

beberapa penelitian menunjukkan hubungan positif antara kepribadian dan penilaian kinerja (Roch, Ayman, Newhouse, & Harris, 2005), sedangkan penelitian lain menunjukkan hubungan negatif (Tyler & Villanova, 2009), dan masih penelitian lain menunjukkan hampir tidak ada hubungan (Tziner, Murphy, & Cleveland, 2002). Oleh karena itu, diperlukan studi-studi lanjutan yang terkait dengan kepribadian penilai dan hasil penilaian mereka untuk menegaskan atau memperjelas hubungan antara variabel-variabel tersebut.

Beberapa model kepribadian telah digunakan dalam beberapa studi atau penelitian untuk mengungkap keterkaitan antara ciri-ciri kepribadian penilai dengan perilaku mereka dalam penilaian. Model kepribadian yang banyak digunakan dalam penelitian ialah *Big Five Personality*. Model kepribadian lima faktor tersebut telah diusulkan sebagai kerangka kerja integratif untuk mempelajari perbedaan kepribadian individu dan merupakan salah satu taksonomi kepribadian yang paling diterima dengan baik dalam literatur (Costa & McCrae, 1992). Menurut sudut pandang ini, struktur kepribadian laten bersifat hierarkis dengan lima faktor pada level tertinggi (Goldberg, 1992; Hough & Ones, 2001). Lima faktor yang dimaksud, yaitu *agreeableness*, *extraversion*, *neuroticism*, *conscientiousness*, dan *openness to experience* (Feist & Feist, 2010).

Agreeableness mencerminkan sifat-sifat mudah percaya, altruisme (mendahulukan kepentingan orang lain), dan menyukai kerja sama (Costa & McCrae, 1992). *Extraversion* berhubungan positif dengan perilaku orang dalam organisasi dan jaringan sosial (Forret & Dougherty, 2001). *Neuroticism* mencerminkan sifat-sifat seseorang yang tenang dan temperamental (Costa & McCrae, 1992) termasuk pula sifat-sifat cemas dan depresi (John et al., 2008). *Conscientiousness* atau kesadaran nurani mencerminkan sifat-sifat teliti dan berorientasi pada prestasi (Costa & McCrae, 1992). *Openness to experience* atau keterbukaan mencerminkan sifat-sifat imajinatif, kreatif, dan rasa ingin tahu (Costa & McCrae, 1992).

Untuk mengetahui ciri-ciri kepribadian seseorang, diperlukan pula instrumen kepribadian yang valid dan reliabel. *Big Five Inventory* (BFI) merupakan salah satu instrumen yang digunakan untuk mengukur kepribadian model lima faktor. Di Indonesia, sudah banyak usaha yang dilakukan untuk

mengadaptasi instrumen kepribadian BFI. Adaptasi instrumen yang dilakukan juga diikuti dengan pengujian validitas dan reliabilitas instrumen tersebut. Namun, berdasarkan pengamatan terbatas, validasi yang dilakukan umumnya masih berdasarkan teori tes klasik. Belum banyak yang melakukan validasi instrumen ini dengan menggunakan teori tes modern, misalnya dengan *Rasch Model*. Oleh karena itu, untuk memperkuat validitas dan reliabilitas instrumen kepribadian dalam bahasa Indonesia, selain dengan model tes klasik, perlu juga dilakukan validasi instrumen dengan menggunakan tes modern, khususnya dengan Rasch Model.

Selain kepribadian penilai, ada faktor lain yang diduga memengaruhi severitas penilai, yaitu faktor latar belakang penilai. Dari beberapa faktor yang menyebabkan severitas tersebut, latar belakang penilai, misalnya usia, gender, pengalaman, dan pelatihan (Pula & Huot, 1993) adalah faktor yang dapat menyebabkan severitas penilai (McNamara, 1996). Penilai yang sudah berpengalaman dan berusia lebih tua cenderung menilai tulisan secara holistik berdasarkan pandangan pribadi atau kesan mereka secara umum, sedangkan penilai yang lebih muda cenderung menilai secara analitis. Pelatihan yang pernah diikuti penilai, baik itu berupa pelatihan kebahasaan maupun pelatihan tentang prosedur penilaian, memungkinkan penilai untuk mencapai konsensus atau kesepakatan mengenai kualitas tulisan karena pelatihan memberikan konsepsi yang jelas tentang kriteria tulisan yang berkualitas.

Penilaian menulis, khususnya jenis tulisan ilmiah, adalah kegiatan atau proses yang sangat penting karena hal ini berkaitan erat dengan masa depan dan kemajuan karier banyak orang. Penilaian menulis akan menentukan pengembangan kualitas tulisan mahasiswa, peneliti, atau akademisi yang berujung pada peningkatan kualitas pendidikan di Indonesia. Oleh karena itu, perlu dilakukan upaya-upaya agar penilaian menulis terjamin keadilan dan objektivitasnya. Upaya yang dapat dilakukan, di antaranya melakukan kajian tentang instrumen penilaian menulis, instrumen untuk mengukur kepribadian, severitas dan konsistensi penilai, keterkaitan antara kepribadian penilai dan latar belakang penilai dalam kegiatan penilaian menulis yang melibatkan penilai

1.2 Pembatasan Masalah

Pembatasan masalah dilakukan untuk menghindari penyimpangan atau pelebaran pokok masalah agar penelitian lebih terarah dan pembahasannya lebih mudah. Instrumen penilaian karya tulis ilmiah pada penelitian ini dibatasi pada instrumen penilaian artikel ilmiah bidang kebahasaan, sedangkan tulisan yang akan dinilai dalam penelitian ini dibatasi pada jenis tulisan artikel ilmiah bidang kebahasaan. Model kepribadian yang digunakan dalam penelitian ini adalah model lima faktor, yaitu *extraversion*, *agreeableness*, *neuroticism*, *conscientiousness*, dan *openness*. Latar belakang penilai dalam penelitian ini dibatasi pada usia, gender, pengalaman menilai, pelatihan di kantor, dan pelatihan di luar kantor.

1.3 Rumusan Masalah

Berdasarkan latar belakang dan pembatasan masalah tersebut, masalah dalam penelitian ini dirumuskan dalam pertanyaan penelitian sebagai berikut.

1. Bagaimanakah analisis instrumen kepribadian *Big Five Inventory* (BFI) berdasarkan teori tes modern (model Rasch)?
2. Bagaimanakah analisis instrumen penilaian artikel ilmiah berdasarkan teori tes modern (*Many Facets Rasch Measurement*)?
3. Bagaimanakah analisis severitas dan konsistensi penilai dalam penilaian artikel ilmiah?
4. Apakah terdapat pengaruh kepribadian penilai (*agreeableness*, *conscientiousness*, *extraversion*, *neuroticism*, *openness*) dan latar belakang penilai (usia, gender, pengalaman menilai, pelatihan di kantor, pelatihan di luar kantor) terhadap severitas penilai?

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut:

1. menganalisis instrumen kepribadian *Big Five Inventory* (BFI) berdasarkan teori tes modern (model Rasch);
2. menganalisis instrumen penilaian artikel ilmiah berdasarkan teori tes modern (*Many Facets Rasch Measurement*);

3. menganalisis severitas dan konsistensi penilai dalam penilaian artikel ilmiah;
4. menyelidiki apakah terdapat pengaruh kepribadian penilai (*agreeableness, conscientiousness, extraversion, neuroticism, openness*) dan latar belakang penilai (usia, gender, pengalaman menilai, pelatihan di kantor, pelatihan di luar kantor) terhadap severitas penilai.

1.5 Signifikansi Penelitian

Penelitian ini diharapkan bermanfaat bagi bidang ilmu pengukuran dan pengembangan instrumen, khususnya instrumen penilaian menulis. Hasil penelitian ini diharapkan dapat menjadi sumber referensi peneliti dan akademisi dalam bidang terkait untuk mengkaji dan mengembangkan penelitian sejenis. Selain itu, penelitian ini juga diharapkan dapat memberikan kontribusi yang bermakna dalam bidang pendidikan bahasa, khususnya dalam kemahiran berbahasa menulis.

Penelitian ini dapat dijadikan rujukan peneliti, mahasiswa, dan khususnya guru bahasa dalam menyusun instrumen penilaian kemahiran berbahasa dalam bentuk rubrik. Manfaat dari aspek akademis lainnya adalah bahwa penelitian ini dapat dijadikan sebagai inspirasi bagi para peneliti untuk melakukan penelitian lanjutan terkait dengan hal-hal yang belum dibahas dalam penelitian ini. Oleh karena itu, penelitian ini terbuka untuk terus dikembangkan dan disempurnakan oleh para peneliti dalam bidang-bidang keilmuan tersebut.

Hasil penelitian ini diharapkan dapat diimplikasikan di dalam dunia penelitian, pengukuran, dan pendidikan. Analisis tes model Rasch dan Rasch Multifaset (MFRM) dapat diaplikasikan dalam menyeleksi penilai yang konsisten dan berkualitas untuk berbagai keperluan, khususnya untuk menyeleksi dan mengevaluasi penilai artikel ilmiah di jurnal kebahasaan dalam lingkungan Badan Pengembangan dan Pembinaan Bahasa, Kemdikbudristek. Secara umum, analisis ini dapat dimanfaatkan untuk berbagai kepentingan, misalnya untuk keperluan penjurian berbagai lomba atau kejuaraan yang menyertakan penilai. Informasi tentang kinerja penilai sangat berarti sebagai umpan balik untuk menentukan penilai atau juri pada kegiatan yang sama berikutnya.

Selain itu, hasil penelitian ini, yaitu pengaruh kepribadian dan latar belakang penilai terhadap severitas penilai, dapat dimanfaatkan untuk menjadi pertimbangan dalam memilih penilai. Ada baiknya, sebelum memilih penilai, dilakukan tes terlebih dahulu tipe kepribadiannya dan diketahui latar belakangnya (usia, gender, pengalaman menilai, dan pelatihan yang pernah diikuti penilai) sehingga penilai dengan tipe kepribadian atau latar belakang tertentu dapat dipilih untuk dilibatkan atau tidak dilibatkan. Pada akhirnya, penelitian ini diharapkan dapat meningkatkan kualitas tulisan akademis di lingkungan universitas, lembaga riset, atau kementerian.

1.6 Kebaruan Penelitian

Ada beberapa studi yang telah dilakukan terkait dengan pengembangan instrumen penilaian menulis di Indonesia. Musdalifa (2015) mengembangkan instrumen penilaian menulis deskripsi, Siregar (2020) mengembangkan instrumen penilaian menulis cerpen, dan Sadri (2021) melakukan pengembangan instrumen menulis puisi. Ketiga studi tersebut mengembangkan instrumen penilaian menulis, tetapi jenis tulisan yang dinilai bukanlah karya tulis ilmiah (KTI). Ada studi pengembangan instrumen penilaian karya tulis ilmiah, yaitu dilakukan oleh Susilo (2019). Keempat studi tentang pengembangan penilaian menulis tersebut berusaha mengembangkan instrumen penilaian menulis berbentuk rubrik. Namun, validasi instrumen pada studi-studi tersebut belum maksimal karena hanya sampai pada tahap validasi isi oleh pakar yang hasilnya dikuantifikasi dengan persentase.

Studi-studi terdahulu belum memvalidasi butir rubrik penilaian berdasarkan parameter-parameter psikometri yang lengkap dan memadai. Di sisi lain, Yamtinah dkk. (2021) melakukan pengembangan instrumen penilaian menulis narasi. Validasi instrumen yang mereka kembangkan itu berupa validasi isi oleh pakar yang hasilnya diolah dengan model *coefficient validity ratio* (CVR) dan reliabilitasnya diuji dengan alpha Cronbach's. Usaha ini dinilai lebih baik jika dibandingkan dengan keempat studi sebelumnya karena sudah melakukan analisis butir penilaian dengan salah satu parameter psikometri. Namun, analisis yang dilakukan Yamtinah dkk. tersebut belum berdasarkan pada parameter psikometri

yang lebih lengkap dan juga belum berdasarkan model tes termutakhir atau tes modern (misalnya model Rasch).

Salim Razi (2015) mengembangkan instrumen rubrik artikel ilmiah yang disebut dengan *Transparent Academic Writing Rubric* (TAWR). Struktur internal rubrik ini divalidasi berdasarkan pendekatan Campbell and Fiske (1959), yaitu validasi konvergensi dan diskriminan. Reliabilitas rubrik ini dibuktikan dengan nilai alpha Cronbach's sebesar 0,89 (Razi, 2015). Instrumen ini ditujukan untuk penilaian artikel ilmiah dalam bahasa Inggris sehingga tidak cocok jika langsung diterapkan untuk menilai artikel ilmiah dalam bahasa Indonesia mengingat adanya perbedaan tataran linguistik antara bahasa Inggris dan bahasa Indonesia. Selain itu, validasi instrumen yang telah dilakukan untuk TAWR ini belum berdasarkan model tes termutakhir atau tes modern.

Studi-studi terdahulu yang terkait dengan pengembangan instrumen kepribadian *Big Five Inventory*, khususnya di Indonesia, antara lain dilakukan oleh Ramdhani (2012), Alawiyah (2015), serta Ahya dan Siaputra (2021). Studi-studi tersebut melakukan pengadaptasian atau penyesuaian instrumen BFI ke dalam bahasa dan budaya Indonesia. Setelah dilakukan adaptasi instrumen, dalam studi-studi tersebut juga dilakukan validasi ulang terhadap instrumen BFI dengan menggunakan model *confirmatory factor analysis* (CFA) yang menghasilkan instrumen BFI berbahasa Indonesia. Namun, studi-studi ini belum menggunakan model analisis tes modern, khususnya Rasch Model.

Studi-studi yang terkait dengan severitas, di antaranya adalah yang dilakukan Mulqueen dkk. (2000), Kaliski dkk. (2012), Amanda Brijmohan (2016), Şahin (2016), dan Primi (2018). Secara umum, penelitian-penelitian tersebut membahas severitas dan konsistensi penilai dalam penilaian prosedur, kreativitas, penilaian sejawat, wawancara dalam konteks suatu pelatihan. Studi lain yang membahas severitas penilai dalam penilaian menulis juga dilakukan oleh Eckes (2009) dan Schaefer (dalam Farrokhi dkk., 2012).

Studi tentang kepribadian penilai yang dikaitkan dengan hasil penilaian juga telah dilakukan Harari dkk. (2014) dan Alaei dkk. (2014). Harari menyelidiki pengaruh tipe kepribadian penilai dalam variasi penilaian kinerja pegawai, sedangkan Alaei dkk. (2014) menyelidiki hubungan antara tipe kepribadian

penilai, metode penilaian analitis/holistis, dan genre tulisan. Hubungan antara kepribadian dan berbagai perilaku manusia yang dijelaskan dengan menggunakan model kepribadian BFI juga telah banyak diteliti, yaitu keterkaitan antara faktor-faktor kepribadian BFI dengan kepuasan kerja (Judge, Heller & Mount, 2002), perilaku keselamatan kerja (Wallace & Vodanovich, 2003), dan gaya kepemimpinan (Bono & Judge, 2004). Penelitian lain juga menyelidiki keterkaitan antara faktor-faktor kepribadian BFI dengan perilaku antisosial (Lee, Ashton & Shin, 2005), motivasi (Judge & Ilies, 2002), dan kesalahan penilai (Bernardin, Cooke, & Villanova, 2000; Yun dkk., 2005).

Berdasarkan penjelasan tersebut diketahui bahwa penelitian-penelitian yang terkait dengan severitas dan faktor-faktor yang melatarbelakanginya masih bersifat parsial atau tidak komprehensif. Ada yang mengkaji severitas penilai, tetapi belum dikaitkan dengan kepribadian penilai. Di sisi lain, ada studi-studi yang telah mengkaji kepribadian penilai yang dikaitkan dengan hasil penilaian, tetapi belum dikaitkan dengan severitas penilai atau latar belakang penilai. Selain itu, ada penelitian yang mengaitkan antara severitas penilai dengan kepribadian penilai, tetapi belum menyorot pada artikel ilmiah sebagai objek penelitiannya.

Oleh karena itu, **kebaruan** atau *novelty* penelitian ini **secara substansial** adalah bahwa penelitian ini lebih komprehensif jika dibandingkan dengan penelitian-penelitian sebelumnya yang bersifat parsial. Dikatakan lebih komprehensif karena penelitian ini berusaha menyelidiki keterkaitan antara severitas penilai dengan gabungan banyak variabel sekaligus, yaitu variabel kepribadian (*agreeableness, conscientiousness, extraversion, neuroticism, dan openness*) dan latar belakang penilai (usia, gender, pelatihan di kantor, dan pelatihan di luar kantor). Kebaruan lainnya adalah penelitian ini mengembangkan instrumen penilaian artikel ilmiah. Instrumen penilaian khusus untuk karya tulis ilmiah berjenis artikel ilmiah masih jarang ditemui di Indonesia. Walaupun ada, instrumen itu tidak tervalidasi atau belum tervalidasi secara memadai.

Kebaruan penelitian ini dari aspek **metodologis**, yaitu bahwa penelitian ini menggunakan model *Many Facets Rasch Measurement (MFRM)* dalam analisis butir soal, penilai, dan kualitas artikel ilmiah. Di Indonesia, model ini masih jarang digunakan untuk memvalidasi instrumen, menganalisis butir penilaian yang

melibatkan penilai, atau menganalisis kualitas artikel ilmiah. Dengan menggunakan model ini akan dihasilkan butir penilaian dengan validitas yang baik. Selain itu, dapat dilakukan analisis karakteristik penilai (severitas dan konsistensinya) dan analisis kualitas artikel ilmiah secara objektif dan komprehensif. Dengan menggunakan model MFRM ini, hasil penilaian menulis, khususnya artikel ilmiah, akan lebih adil dan objektif. Selain itu, kebaruan metodologis penelitian ini adalah validasi instrumen BFI dengan menggunakan model CFA (*confirmatory factor analysis*) dan model Rasch. Dengan demikian, validasi instrumen-instrumennya dilakukan secara maksimal.

Adapun *state of the art* atau kedudukan penelitian ini adalah bahwa penelitian ini adalah penelitian pengembangan dari penelitian-penelitian yang telah dilakukan sebelumnya. Instrumen yang dikembangkan dalam penelitian ini adalah instrumen penilaian karya tulis ilmiah *Transparent Academic Writing Rubric* (TAWR), yaitu rubrik penilaian karya tulis ilmiah yang dikembangkan oleh Salim Razi (2015). Pada penelitian ini, rubrik TAWR akan divalidasi ulang dengan model *Many Facets Rasch Measurement* (MFRM). Selain itu, rubrik penilaian menulis ini akan dikembangkan dengan menambahkan panduan penilaian berupa penjelasan dan indikator untuk tiap-tiap butir penilaian artikel ilmiah. Instrumen yang dikembangkan berikutnya adalah instrumen *Big Five Inventory* (Goldberg, 1981). Instrumen BFI itu akan diadaptasi dan divalidasi ulang dengan menggunakan model Rasch. Pengembangan penelitian lainnya adalah terkait dengan severitas penilai dalam penilaian. Pada penelitian ini severitas penilai akan dikaitkan dengan kepribadian dan latar belakang penilai.