

PENGIMPLEMENTASIAN LIMA PRINSIP BROWN DALAM MENDESAIN ASSESMEN PEMBELAJARAN BAHASA INGGRIS

Muhammad Chanif Tamimi Raditya Wandana

Universitas Negeri Malang, Jalan Semarang No.5, Sumbersari, Kec. Lowokwaru,
Kota Malang, Provinsi Jawa Timur. Kode Pos: 65145
E_mail: raditya.wandana@gmail.com

Abstrak: Studi ini mengeksplorasi artikel oleh Kosimov (2022) dengan judul “Pentingnya memilih alat penilaian yang tepat dalam pengajaran bahasa dan dampaknya terhadap penguasaan bahasa kedua.”. Penelitian ini bertujuan untuk memahami bagaimana pengimplementasian lima prinsip Brown dalam assessmen Bahasa Inggris. Hasilnya menunjukkan bahwa peneliti tidak mengikuti metode penelitian yang jelas dengan ambiguitas dalam menangani penelitian evaluasi dan pengembangan. Namun, penelitian ini tidak berfokus pada evaluasi atau mengikuti metode penelitian dan pengembangan umum yang sering digunakan dalam mendesain media. Masalah lainnya adalah kurangnya bukti empiris untuk mendukung klaim dan perubahan apa pun yang dilakukan pada alat penilaian dalam penelitian. Satu-satunya landasan teoretis utama untuk penelitian ini adalah lima prinsip penilaian bahasa Brown (2004). Namun, peneliti tidak dapat melakukan analisis yang konsisten dan menyeluruh dengan menggunakan teori tersebut.

Kata Kunci: Penilaian Bahasa, Pengajaran Bahasa Inggris, Prinsip Penilaian

THE IMPLEMENTATION OF BROWN’S FIVE PRINCIPLE IN EVALUATING AND DESIGNING A LANGUAGE ASSESSMENT

Abstract: This study explores the article by Kosimov (2022) under the title of “The importance of choosing appropriate assessment tools in language teaching and its impact to second language acquisition.”. The purpose of this study is to analyze the implementation of Brown’s five principle in English language assessment. The result shows that the researcher did not follow any clear research method with the ambiguity of tackling both evaluation and development research. However, this research is neither focuses on the evaluation nor follow the common research and development method that often used in designing media. The other issues would the lack of empirical evidence to support any of the claims and changes made to the assessment tools in the research. The only major theoretical foundation for this research is Brown’s (2004) five principles of language assessment. However, the researcher was unable to conduct a consistent and thorough analysis using the theory.

Keywords: Language Assessment, English Language Teaching, Principle of Assessment,

PENDAHULUAN

Pentingnya asesmen sebagai bagian dari proses pembelajaran bahasa telah dinyatakan melalui banyak gambaran

seperti kumpulan sistem untuk meningkatkan perkembangan siswa (Banta & Palomba, 2015), jembatan antara proses belajar mengajar (William,

2013), serta fungsinya sebagai metode untuk memeriksa kemampuan (Brown & Abeywickrama, 2018) dan kemampuan siswa. Untuk membantu guru mengidentifikasi kelemahan siswa (Llorente & Morant, 2011). Hanya itu yang diperlukan untuk meringkas seluruh poin pengantar dalam penelitian yang dilakukan oleh Kosimov (2022) dengan judul Pentingnya Memilih Alat Asesmen yang Tepat dalam Pengajaran Bahasa dan Dampaknya terhadap Pemerolehan Bahasa Kedua. Penjelasan tentang pentingnya penilaian belum mencapai poin argumentatif untuk membenarkan kebutuhan untuk melakukan penelitian apapun. Tidak ada elaborasi terhadap semua definisi dan deskripsi dari para peneliti yang disebutkan sebelumnya untuk mengikat semua sumber yang dikutip sebagai satu pernyataan tesis yang jelas.

Lebih-lebih lagi, Kosimov (2022) kemudian langsung menyebutkan bahwa penelitian yang dilakukannya akan mengevaluasi kekurangan sekaligus memodifikasi ulangan tengah semester yang digunakan di kalangan mahasiswa di Sekolah Tinggi Profesi Pariwisata dan Jasa Andijan menurut Brown's (2004) lima prinsip penilaian.

Evaluasi Penilaian dapat dilakukan dengan berbagai cara. Hal ini dapat

dilakukan melalui studi kasus dengan wawancara semi terstruktur untuk menggali pandangan siswa terhadap suatu penilaian (Zuhriyah & Pratolo, 2020). Pilihan lain adalah melakukan penelitian kualitatif deskriptif untuk membuat analisis yang komprehensif tentang sifat penilaian tertentu dan pengalaman sekelompok siswa tertentu dalam menggunakan penilaian (Sari et al., 2020). Mengevaluasi penilaian juga dapat dilakukan melalui penelitian kuasi-eksperimental untuk membedakan hasil antara kelompok eksperimen dan kelompok kontrol terhadap penilaian (Chien et al., 2020).

Selain itu, memodifikasi penilaian akan menciptakan alat baru dan lebih baik untuk menilai siswa yang memimpin orang percaya bahwa penelitian dan pengembangan bersama dengan analisis kebutuhan akan menjadi metode yang tepat untuk membuat alat penilaian yang valid dan handal (Fathurohman & Cahyaningsih, 2021; Nyoman Padmadewi & Putu Artini, 2019; Pranoto & Suprayogi, 2020). Penilaian yang baru dikembangkan dapat didasarkan pada [masukkan penelitian sebelumnya]. Selain itu, kurangnya metode yang dinyatakan dengan jelas juga berkontribusi pada

kebingungan terhadap ruang lingkup metode penelitian.

METODE PENELITIAN

Penelitian mengkaji dan melakukan *critical review* terhadap karya ilmiah Kosimov (2022) dengan judul "*The importance of choosing appropriate assessment tools in language teaching and its impact to second language acquisition.*".

Penelitian ini mengidentifikasi kelebihan dan kekurangan dari karya ilmiah tersebut, mengungkapkan kontribusi karya tersebut terhadap pengembangan disiplin ilmu, serta memberikan saran dan rekomendasi bagi penulis dan para pembaca terkait dengan peningkatan kualitas karya ilmiah yang dihasilkan di Indonesia.

Melalui penelitian ini, diharapkan dapat meningkatkan pemahaman dan kesadaran para penulis dan pembaca tentang pentingnya melakukan *critical review* terhadap karya ilmiah yang diterbitkan, sehingga dapat memperbaiki kualitas dan relevansi penelitian serta meningkatkan kontribusinya terhadap pengembangan ilmu pengetahuan di Indonesia.

HASIL PENELITIAN

Mendesain Asessmen

Menurut Kosimov (2022), anak-anak di Uzbekistan mulai belajar bahasa Inggris sejak usia muda dan para guru sering menggunakan metode Grammar-Translation daripada metode modern lainnya seperti Pengajaran Bahasa Komunikatif. Pemerintah Uzbekistan juga mendorong siswa untuk memiliki sertifikat kemahiran bahasa yang diakui secara internasional dengan memberikan manfaat tambahan berupa menerima 50 poin dalam bahasa asing dan tidak ada persyaratan untuk mengikuti tes masuk bahasa asing yang diadakan dua kali setahun. Ini menunjukkan betapa pentingnya tingkat kemahiran bahasa Inggris siswa di Uzbekistan dan juga pentingnya alat penilaian dalam pengajaran bahasa.

Sekolah Tinggi Pariwisata dan Layanan Profesional Andijan berfokus pada pelatihan siswa untuk industri. Salah satu bahasa asing terpopuler di Uzbekistan adalah bahasa Inggris, yang diajarkan di institusi empat kali seminggu menggunakan sejumlah teknik, termasuk CLT. Alat bantu pengajaran teknis disediakan di kelas bahasa Inggris, dan latihan kontrol dilakukan setiap tiga hingga empat pelajaran. Pedoman Kementerian Pendidikan Umum menyatakan bahwa siswa harus memiliki tingkat

kecakapan bahasa Inggris A2+. Siswa yang lulus tes masuk dan ditempatkan di salah satu dari tiga kelompok sesuai dengan kedudukan akademik mereka diterima di perguruan tinggi. Inisiatif yang diambil oleh Kementerian Pendidikan Umum Uzbekistan untuk meningkatkan pengajaran bahasa Inggris di negara tersebut. Guru di tempat-tempat seperti Sekolah Tinggi Pariwisata dan Layanan Profesional Andijan menggunakan teknik mutakhir dan evaluasi formatif untuk membantu siswa dengan keterampilan tata bahasa dan komunikasi mereka. Program ini sangat menekankan pada pengukuran kedua keterampilan tersebut karena diharapkan para lulusan mampu berkomunikasi dengan penutur asli bahasa Inggris. Guru dapat merencanakan kelas, memeriksa tingkat pemahaman siswa, dan menghasilkan metode pengajaran baru dengan menggunakan penilaian formatif. Menurut Brown, instruktur harus menggunakan bukti spesifik. Pada akhir semester atau tahun, evaluasi sumatif digunakan untuk mengevaluasi kinerja siswa secara keseluruhan.

Evaluasi sumatif yang dibahas oleh Kosimov (2022) adalah ujian tengah semester bagi mahasiswa semester pertama yang terdiri dari ujian tertulis

berbasis tata bahasa dan ujian berbicara. Peneliti lebih lanjut menjelaskan bahwa siswa tidak mengalami kesulitan dengan tes berbasis tata bahasa karena mudah untuk mencontek dan menemukan jawaban untuk tes tersebut. Namun, peneliti tidak merinci dari mana klaim tersebut berasal. Tidak ada bukti pendukung yang secara eksplisit mengarah pada pembuatan klaim yang disebutkan sebelumnya. Lanjut ke tes berbicara, siswa dikatakan mengalami kesulitan dalam menjawab soal karena rumitnya soal untuk siswa baru. Tidak ada penjelasan rinci tentang apa yang didefinisikan sebagai mahasiswa baru, tetapi karena penelitian dilakukan dengan mahasiswa semester pertama, dapat diasumsikan bahwa kedua istilah tersebut agak terkait. Untuk lebih menjelaskan partisipan penelitian, peneliti menyebutkan satu kelompok yang terdiri dari 24 mahasiswa dengan penyorotan kepada satu mahasiswa bernama Amurbek. Tidak ada alasan ilmiah khusus mengapa Amurbek dipilih juga untuk peran yang tidak jelas dan tidak ditentukan dalam grup selain dia adalah yang terbaik dari grup. Asumsi yang mungkin adalah bahwa bahkan siswa terbaik pun mengalami kesulitan dalam melakukan penilaian, ini akan

menunjukkan bahwa ada beberapa masalah dengan penilaian itu sendiri. Sayangnya, peneliti tidak memberikan latar belakang teoretis apa pun untuk mendukung klaim tersebut dengan asumsi itulah yang ingin dikatakan oleh peneliti.

Kurangnya instrumen yang valid atau instrumen apa pun dalam hal ini

akan menjadi poin selanjutnya yang akan dibahas dalam ulasan ini. Setelah menyebutkan kesulitan-kesulitan yang dihadapi siswa, Kosimov (2022) kemudian menganalisis dan mengkritisi alat penilaian dengan secara eksplisit memberikan beberapa contoh dari pertanyaan lisan. Pertanyaannya adalah sebagai berikut:

Tabel 1. Contoh Pertanyaan Assesmen Kosimov (2022)

No	Pertanyaan
1	“Apa tujuan dari pariwisata berkelanjutan?”
2	“Apa faktor spesifik yang mempengaruhi perkembangan pariwisata di Uzbekistan?”
3	“Apakah Anda mendukung keterbukaan turis yang diumumkan oleh beberapa negara, mengetahui bahwa dunia belum menemukan pengobatan atau vaksin yang efektif?”

Namun, alih-alih menganalisis pertanyaan melalui instrumen, peneliti hanya menyatakan bahwa tidak ada pertimbangan spesifikasi tes yang diambil oleh guru dan tidak tepat untuk diberikan kepada siswa karena penilaian mengandung informasi yang tidak relevan terkait dengan kemampuan siswa. tingkat. Dengan demikian, peneliti memang mengacu pada lima prinsip Brown (2004) dan beberapa penelitian sebelumnya. Namun, semua latar belakang teori yang diberikan hanya berdasarkan gambaran umum tentang bagaimana penilaian yang ideal seharusnya. Peneliti harus mencerminkan

penelitian analisis kebutuhan karena menganalisis beberapa aspek siswa seperti kurikulum, tingkat siswa, kebutuhan dan pendapat terhadap kursus bahasa Inggris termasuk penilaian sehingga ada bukti empiris untuk mendukung semua klaim yang disebutkan sebelumnya (Asrifan et al., 2020). Sebaliknya, peneliti mengambil sendiri untuk memberikan elaborasi terhadap isu-isu penilaian. Pertama, tes tata bahasa tertulis tersebut dirasa terlalu mudah dan seharusnya dirancang agar lebih bermanfaat (Terrell, 1991) dan efektif karena sangat penting untuk mengetahui struktur yang digunakan

dalam komunikasi (Savignon, 1991). Namun, selain mengutip dua sumber dari tahun 90-an, peneliti tidak memberikan data dan penjelasan lebih lanjut. Bagaimana tepatnya penilaian tata bahasa harus dilakukan.

Kedua, untuk tes berbicara, peneliti menyebutkan empat dari lima prinsip penilaian yang meliputi kepraktisan, validitas, reliabilitas dan keaslian. Untuk kepraktisan, peneliti menjelaskan bahwa karena penilaian tidak memberikan instruksi yang jelas sehingga guru tidak mempertimbangkan apakah dapat dilakukan dalam kerangka waktu. Namun, peneliti tidak menyebutkan instruksi yang tidak jelas tersebut atau bahkan batas waktu saat ini dan bukti lain untuk membenarkan klaim tersebut. Untuk validitas, peneliti memberikan bukti yang jauh lebih baik untuk mendukung klaim bahwa hasil tes tidak dapat mengukur pengetahuan tata bahasa siswa dengan secara eksplisit menyebutkan fakta bahwa penilaian hanya memiliki 4 dari 5 tenses dan harus mencakup topik yang telah dipelajari siswa. Namun, akan lebih baik jika peneliti menyebutkan apa saja 4 tenses dan topik yang telah dipelajari siswa. Untuk reliabilitas, peneliti menunjukkan bahwa asesmen tidak sesuai dengan

kemampuan peserta tes karena berbicara adalah keterampilan yang cukup menantang untuk dinilai sehingga asesor harus mempertimbangkan kebutuhan dan karakteristik siswa sebelum merancang penilaian. Dengan demikian, peneliti mengelaborasi seperti apa karakteristik dan kebutuhan peserta didik serta apa yang sesuai dengan kemampuan peserta tes. Untuk pertama kalinya, peneliti setuju bahwa penilaian tersebut telah memenuhi kriteria keaslian karena pertanyaannya terkait dengan keadaan aktual terkait masalah pariwisata. Namun, keaslian seperti itu disebutkan membuat pertanyaan lebih sulit dijawab.

Kosimov (2022) kemudian memasukkan semua analisis di atas ke dalam kelebihan dan kekurangan penilaian. Penilaian digambarkan sebagai sangat kritis dan akurat untuk konteks kehidupan nyata. Peneliti juga menyebutkan bahwa asesmen meningkatkan pengetahuan tata bahasa dan meningkatkan akurasi komunikasi meskipun tes tata bahasa dianggap terlalu mudah dan seharusnya lebih efektif. Untuk kekurangannya, sama saja dengan analisis sebelumnya bahwa asesmen tidak dapat mengukur kemampuan siswa secara akurat, tidak terlalu memperhati-

kan tingkat dan kebutuhan siswa dan yang terpenting asesor tidak mempertimbangkan lima prinsip Brown (2004).

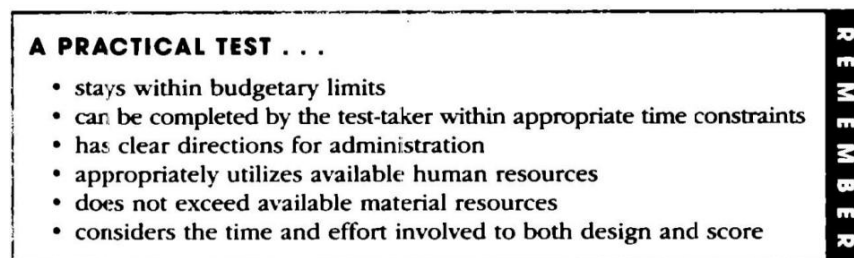
PEMBAHASAN

Lima Prinsip Penilaian Brown (2004)

Lima prinsip penilaian bahasa yang dikemukakan oleh Brown (2004) meliputi kepraktisan, reliabilitas, validitas, keaslian dan washback. Kelima prinsip tersebut berkontribusi pada aspek penilaian yang bergantung pada konteks yang berbeda dengan tujuan menciptakan penilaian bahasa yang sesuai. Namun, prinsip ini juga dapat diterapkan pada semua jenis penilaian.

Kepraktisan

Yang pertama adalah kepraktisan. Ini mengacu pada logistik melakukan penilaian. Ini mencakup hal-hal seperti biaya, waktu, penyelenggaraan tes, kemudahan pemahaman dan penilaian. Untuk menganalisis kepraktisan suatu alat asesmen perlu dipertimbangkan pengumpulan data dari proses pembuatan asesmen, proses saat pelaksanaan asesmen dan proses penilaian atau setelah asesmen dilakukan. Brown (2004) mengemukakan kepraktisan yang ideal dari sebuah tes sebagai berikut.



Gambar 1. Gambaran Brown tentang tes praktek

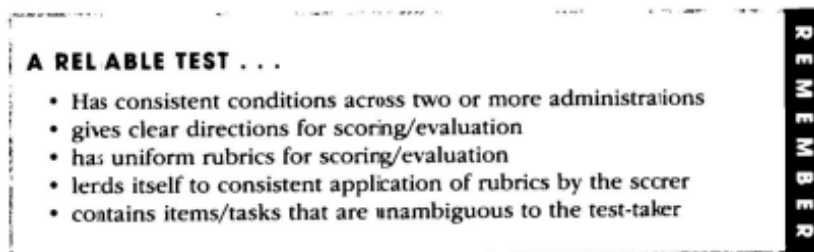
Kriteria di atas dapat digunakan sebagai instrumen untuk mengidentifikasi apakah suatu tes dianggap praktis atau tidak. Tidak hanya itu Kosimov (2022) tidak menggunakan instrumen untuk menganalisis kepraktisan penilaian, juga tidak ada data empiris yang dapat

menyerupai kriteria kepraktisan yang mungkin.

Reliabilitas

Karena penilaian yang dianalisis oleh Kosimov (2022) merupakan tes tengah semester yang artinya berupa tes sumatif dan standar. Keandalan memastikan bahwa tes standar memiliki standar.

Sehingga meskipun tes penilaian dilakukan dengan asesor, siswa, waktu dan tempat yang berbeda, namun hasilnya tetap konsisten. Brown (2004) mengemukakan reliabilitas ideal suatu tes sebagai berikut.



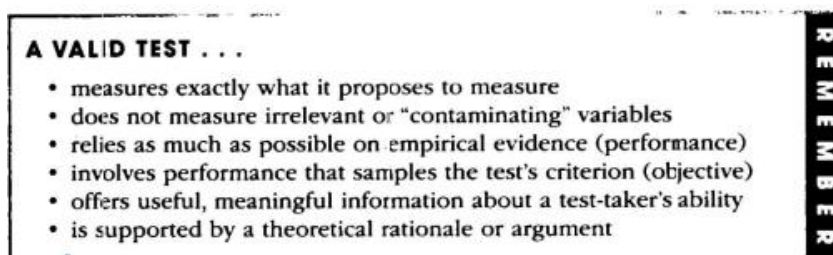
Gambar 2. Gambar 1. Gambaran Brown tentang tes reliabel

Kemungkinan data yang seharusnya disediakan oleh Kosimov (2022) adalah tentang instruksi yang diberikan dalam penilaian dan konsistensi antara penilai yang berbeda. Sebaliknya, peneliti berfokus untuk menunjukkan bahwa penilaian tersebut tidak sesuai kebutuhan dan karakteristik siswa.

Keabsahan

Sementara analisis Kosimov (2022) terhadap validitas asesmen telah

menyebutkan tentang pentingnya menilai hanya topik yang telah dipelajari siswa, prinsip validitas memiliki penjelasan yang jauh lebih rinci. Pertama, penilaian harus dapat menilai apa yang seharusnya dinilai dan seperti yang disebutkan oleh peneliti, ujian tengah semester tidak dapat mengukur pengetahuan tata bahasa siswa sehingga menjadikannya bukan alat penilaian yang valid. Lebih lanjut, Brown (2004) mengemukakan validitas ideal suatu tes sebagai berikut.



Gambar 3. Gambaran Brown tentang tes yang valid

Namun, cukup penting bahwa validitas sangat bergantung pada rasional dan argumentasi teoretis, seperti yang

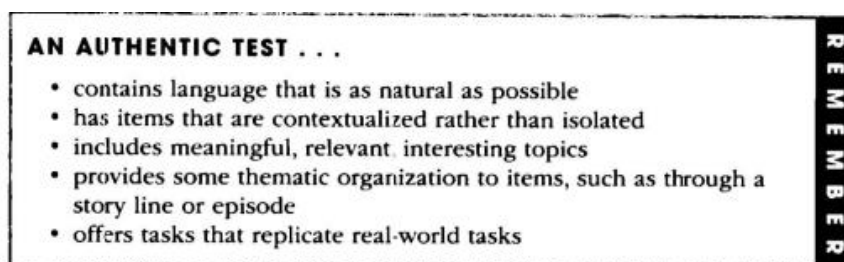
tertera pada gambar di atas. Bukti penting untuk membuat penilaian menjadi valid. Hal ini juga dapat digunakan untuk

menunjukkan validitas penilaian. Peneliti tidak menyebutkan secara eksplisit bukti rasionalitas yang mempengaruhi validitas UTS selain hanya menyatakan bahwa tes menggunakan 4 dari 5 tenses. Tidak ada konteks lebih lanjut yang kami berikan

Keaslian

Penilaian harus menyerupai tugas atau situasi dunia nyata sehingga siswa memiliki pemahaman yang lebih baik tentang apa yang mereka pelajari. Namun, Kosimov (2022) berpendapat bahwa prinsip ini meningkatkan kesulitan penilaian lisan dari ujian tengah semester.

Dengan demikian, bukan berarti keaslian akan selalu menambah kesulitan. Bagi beberapa orang, keaslian memiliki dampak sebaliknya dalam membuat penilaian lebih mudah dan lebih bermanfaat bagi siswa untuk mengembangkan keterampilan yang berguna karena mencerminkan apa yang biasanya digunakan siswa dalam kehidupan nyata (Drezner, 2022; Jopp, 2020; Sokhanvar et al., 2021). Untuk memastikan keaslian akan menambah lebih banyak penilaian dan tidak hanya meningkatkan kesulitan mereka, Brown (2004) menyarankan keaslian tes yang ideal sebagai berikut.



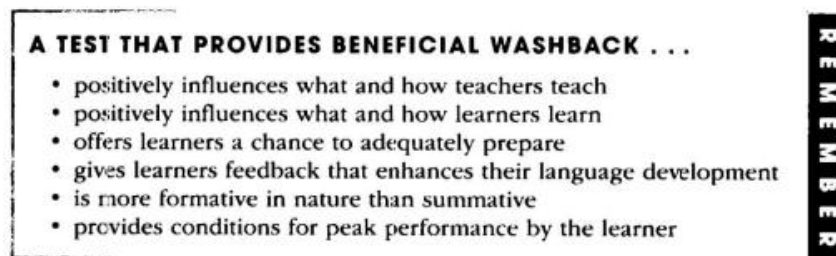
Dikatakan bahwa otentik tidak hanya menyerupai kehidupan nyata tetapi juga bermakna, relevan dan menarik. Tes tengah semester yang dianalisis oleh peneliti mungkin hanya berfokus pada akurasi kehidupan nyata daripada kriteria lainnya. Oleh karena itu, saya akan jauh lebih baik bagi peneliti untuk menganalisis keaslian suatu tes di luar kemiripannya dengan situasi dunia nyata.

Dampak Berkelanjutan (Washback)

Setelah semua prinsip yang disebutkan sebelumnya di atas, penting juga untuk mengukur dampak penilaian terhadap siswa. Kosimov (2022) tidak membahas prinsip ini, meskipun peneliti menyebutkan pengaruhnya terhadap mahasiswa baru atau semester pertama. Apalagi dalam ujian tengah semester yang dianggap sebagai ujian standar yang umumnya diasosiasikan dengan

washback negatif seperti siswa merasa stres dan menjejali semua materi yang dibutuhkan untuk ujian. Brown (2004)

menyarankan *washback* yang bermanfaat dari tes sebagai berikut.



Gambar 5. Gambaran Brown tentang *washback* yang bermanfaat dari suatu tes

Washback positif adalah pertimbangan yang sangat penting dalam merancang penilaian. Oleh karena itu, mengecualikan satu aspek dari prinsip akan menyebabkan penilaian baru dan lebih baik seperti yang ingin dilakukan peneliti menjadi rentan terhadap pencucian balik negatif.

Merancang Penilaian Baru

Sebagian besar studi yang berfokus pada evaluasi alat penilaian sebagian besar menggambarkannya melalui data empiris seperti sudut pandang siswa dan guru, kinerja siswa, kepraktisan dan aspek lain yang menyerupai lima prinsip Brown (2004) (Chien et al., 2020; Sari et al., 2020; Zuhriyah & Pratolo, 2020). Namun, sebagian besar hanya berfokus pada menunjukkan fakta apakah alat penilaian tertentu cocok untuk tujuan tertentu termasuk kelebihan dan kekura-

ngannya. Mereka jarang juga menyertakan produk akhir yang dibuat dengan baik dari alat penilaian yang baru dirancang seperti yang biasanya sedang dalam penelitian dan pengembangan (Fathurohman & Cahyaningsih, 2021; Nyoman Padmadewi & Putu Artini, 2019). Namun, Kosimov (2022) mencoba memodifikasi tes tengah semester untuk meningkatkan penilaian agar lebih sesuai dengan lima prinsip penilaian bahasa Brown (2004). Namun karena penelitian ini tidak mengikuti metode penelitian dan pengembangan yang biasa, maka banyak terjadi perubahan dalam modifikasi ujian tengah semester terpilih yang kurang memiliki data empiris dan landasan teori. Pertama, peneliti menyebutkan bahwa alat penilaian kurang memiliki instruksi yang jelas. Kemudian, peneliti hanya menambahkan instruksi, seperti berikut ini.

N	Tenses	Time	Auxiliary verbs	Sentence
1	Present simple			I do not like dogs
2			Am, is, are	I am reading a book
3	Present perfect Continuous	Past up to now		
4		Now		I am not sleeping
5	Future Simple			I will do it.

Gambar 6. Tes asli (kiri) dan instruksi yang dimodifikasi (kanan)

Meskipun instruksinya jelas dan memenuhi tujuan penilaian, tidak ada data empiris yang mendukung atau mengkonfirmasi apakah instruksi itu mudah dipahami oleh siswa dan secara akurat mengikuti makna yang dimaksud dari guru. Dalam penelitian dan pengembangan berbasis metode ADDIE yang merupakan singkatan dari Analyse,

Design, Develop, Implement, and Evaluate (Furwana et al., 2021), proses konfirmasi pilihan desain termasuk dalam evaluasi.

Begitu pula dengan modifikasi berikutnya, di mana peneliti mengubah pertanyaan berbicara menjadi versi yang lebih mengelaborasi dari penilaian awal. Contohnya sebagai berikut.

Speaking questions.

Card 1

1. "What is the aim of sustainable tourism?"
2. "What are the specific factors affecting tourism development in Uzbekistan?"
3. Most local hospitality and tourism enterprises hardly engage in corporate social responsibility programs within their location. Is this entirely proper?

Card 1.

1. What are the advantages and disadvantages of the tourism industry?
2. What are the main factors affecting tourism?
3. Is it possible to book a holiday to a seaside resort on the other side of the world from home?

Gambar 7. Pertanyaan berbicara asli (atas) dan pertanyaan yang dimodifikasi (bawah)

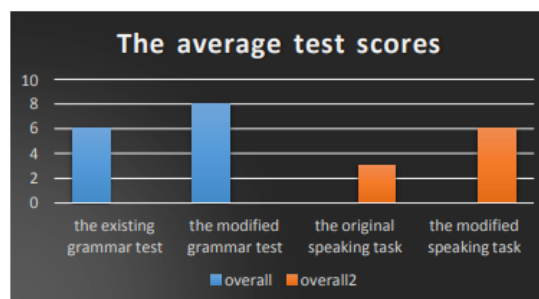
Karena pertanyaan yang dimodifikasi cenderung lebih detail, hal itu akan mengganggu tujuan yang ingin dicapai dari materi pelajaran itu sendiri. Untuk memberikan dasar yang kuat terhadap perubahan tersebut, ada beberapa bukti empiris yang dapat diberikan seperti

analisis kurikulum, wawancara guru dan perspektif siswa terhadap pertanyaan tersebut. Semua bukti itu akan dimasukkan dalam metode penelitian dan pengembangan yang tepat, di bagian analisis kebutuhan (Asrifan et al., 2020).

Kosimov (2022) juga menyebutkan bahwa tes tata bahasa tertulis dari ujian tengah semester diubah dari berbasis kertas menjadi tes online. Kali ini peneliti melakukan survey pendapat siswa untuk melihat perspektif siswa terhadap perubahan tersebut. Hasilnya menunjukkan bahwa baik hasil tes maupun persepsi siswa semuanya positif. Namun, peneliti gagal untuk mempertimbangkan pengaruh perubahan media

penilaian dengan tujuan awal dari tes tersebut. Akan lebih baik jika ada bukti bahwa mengubah media tidak mempengaruhi tujuan penilaian yang diinginkan dari sudut pandang guru dan juga siswa.

Pada akhirnya, penelitian ini menunjukkan peningkatan pada tes grammar tertulis dan tes berbicara seperti yang ditunjukkan pada grafik di bawah ini.



Gambar 8. Hasil nilai tes

Namun, banyak keputusan dalam memodifikasi penilaian yang kurang dalam hal bukti empiris dan landasan teoritis. Selanjutnya, peneliti kemudian membuat daftar keuntungan dari tes yang dimodifikasi dengan beberapa kemiripan dengan tujuan awal pembuatan penilaian berdasarkan lima prinsip Brown (2004). Keuntungan pertama yang dinyatakan adalah tes yang dimodifikasi memiliki instruksi yang jelas, yang peneliti sebutkan sebagai peningkatan ke arah

kepraktisan, validitas dan backwash tes yang mengejutkan. Namun dua keuntungan berikutnya, yang mencakup tes diagnostik dan rubrik yang jelas cukup ambigu tanpa menyebutkan latar belakang teoretis apa pun. Kelebihan yang terakhir dinyatakan berdasarkan pada kehidupan nyata, dimana peneliti memang menyebutkan prinsip otentisitas penilaian bahasa. Jadi dari lima prinsip, peneliti menyebutkan empat di antaranya. Namun, penjelasan-penjelasan

tersebut hanyalah gambaran umum dari keempat prinsip tersebut tanpa korelasi kontekstual dan bukti dari tes penilaian yang sebenarnya, dalam hal ini adalah tes tengah semester.

SIMPULAN

Dari awal penelitian, banyak penjelasan tentang apa yang akan dilakukan penelitian ini termasuk judulnya cukup menyesatkan dan tidak jelas. Penelitian ini tidak secara eksplisit menyebutkan metode sekaligus melakukan penelitian dan pengembangan berbasis evaluasi dan penelitian. Banyak penelitian terdahulu dan latar belakang teoritis yang dikemukakan hanya sebagai gambaran umum yang tidak ada korelasinya dengan konteks penelitian beserta argumentasi yang akan disajikan sebagai pernyataan tesis. Sementara hasil menunjukkan dampak positif dari penelitian ini, kurangnya bukti empiris menunjukkan bahwa validitas penelitian ini dapat diperdebatkan karena mungkin ada perubahan yang disebabkan oleh variabel yang sama sekali berbeda yang tidak dapat dipertanggungjawabkan. Satu-satunya landasan teoretis yang berulang kali disebutkan dalam penelitian ini adalah lima prinsip penilaian bahasa Brown

(2004). Jika demikian, peneliti gagal melakukan analisis dan pengembangan penilaian yang konsisten dan mendalam berdasarkan teori yang diajukan.

DAFTAR RUJUKAN:

- Asrifan, A., Vargheese, K., & Amir, M. (2020). ESP Course Design: Analisis Kebutuhan Jurusan Pariwisata di SMK Indonesia. *Journal of Advanced English Studies*, 3 (2), 69-77. <http://sastra.unifa.ac.id/journal/index.php/jes/index>
- Banta, TW, & Palomba, CA (2015). *Esensi Penilaian: Perencanaan, Pelaksanaan, Peningkatan Penilaian di Perguruan Tinggi* (2nd ed.). Jossey-Bass.
- Coklat, HD (2004). *Prinsip Penilaian Bahasa dan Praktik Kelas* (edisi pertama). Longman.
- Coklat, HD, & Abeywickrama, P. (2018). *Prinsip Penilaian Bahasa dan Praktek Kelas* (edisi ke-3). Pearson.
- Chien, SY, Hwang, GJ, & Jong, MSY (2020). Efek penilaian teman sebaya dalam konteks realitas virtual berbasis video sferis pada kinerja berbahasa Inggris dan persepsi pembelajaran siswa EFL. *Komputer dan Pendidikan*, 146. <https://doi.org/10.1016/j.compedu.2019.103751>
- Drezner, C. (2022). *Menyeimbangkan Pembelajaran Berbasis Bermain dengan Penilaian Otentik dalam Pendidikan Anak Usia Dini: Studi Kasus Perspektif Profesional Anak Usia Dini tentang Portofolio Digital* [Disertasi]. Universitas Northcentral.
- Fathurohman, I., & Cahyaningsih, RD (2021). *Pengembangan Instrumen Asesmen Autentik untuk*

- Meningkatkan Kapasitas Ekstensif Perspektif Metakognitif. *Jurnal Penilaian Asia dalam Pengajaran dan Pembelajaran*, 11 (1), 13-21. <https://doi.org/10.37134/ajatel.vol11.1.2.2021>
- Furwana, D., Palopo, I., & Syam, AT (2021). "Listening is Hard": Model ADDIE pada Pengembangan Lembar Kerja Listening Bahasa Inggris. *Jurnal Bahasa dan Sastra*, 16 (1). <http://journal.unnes.ac.id>
- Jopp, R. (2020). Studi kasus inisiatif pembelajaran yang ditingkatkan teknologi yang mendukung penilaian otentik. *Mengajar di Perguruan Tinggi*, 25 (8), 942-958. <https://doi.org/10.1080/13562517.2019.1613637>
- Kosimov, A. (2022). Pentingnya Memilih Alat Asesmen yang Tepat dalam Pengajaran Bahasa dan Dampaknya terhadap Penguasaan Bahasa Kedua. *Tampilan Inggris*, 7 (1), 16-32. <https://doi.org/10.5281/zenodo.5847595>
- Llorente, R., & Morant, M. (2011). Penambahan Data di Perguruan Tinggi. Di K. Funatsu (Ed.), *Teknologi Fundamental Baru dalam Penambahan Data*. InTech. <https://doi.org/10.5772/13320>
- Nyoman Padmadewi, N., & Putu Artini, L. (2019). Instrumen Asesmen Peningkatan Keterampilan Pengajaran Bahasa Inggris melalui Microteaching di Indonesia. Artikel di *Jurnal EFL Asia*. <https://www.researchgate.net/publication/332173981>
- Pranoto, BE, & Suprayogi, S. (2020). Analisis Kebutuhan ESP Bagi Siswa Pendidikan Jasmani di Indonesia. *Premis: Jurnal Pendidikan Bahasa Inggris*, 9 (1), 94. <https://doi.org/10.24127/pj.v9i1.2274>
- Sari, ABP, Iswahyuni, D., Rejeki, S., & Sutanto, S. (2020). Formulir Google sebagai Alat Penilaian EFL: Fitur dan Keterbatasan Positif. *Premis: Jurnal Pendidikan Bahasa Inggris*, 9 (2), 231-250. <https://doi.org/10.24127/pj.v9i2.3037>
- Savignon, SJ (1991). *Pengajaran Bahasa Komunikatif: Canggih*. *Guru Bahasa Inggris untuk Penutur Bahasa Lain*, 25 (2), 261-277. <https://about.jstor.org/terms>
- Sokhanvar, Z., Salehi, K., & Sokhanvar, F. (2021). Keuntungan dari penilaian otentik untuk meningkatkan pengalaman belajar dan keterampilan kerja mahasiswa pendidikan tinggi: Tinjauan literatur yang sistematis. *Kajian Evaluasi Pendidikan*, 70. <https://doi.org/10.1016/j.stueduc.2021.101030>
- Terrell, TD (1991). Peran Instruksi Tata Bahasa dalam Pendekatan Komunikatif. *Jurnal Bahasa Modern*, 75 (1), 52-63. <https://www.jstor.org/stable/329834>
- William, D. (2013). Penilaian: Jembatan antara Mengajar dan Belajar. *Suara dari Tengah*, 21 (2), 15-20. www.readwritethink.org
- Zuhriyah, S., & Pratolo, BW (2020). Menggali pandangan siswa dalam penggunaan quizizz sebagai alat penilaian dalam bahasa Inggris sebagai bahasa asing (EFL) kelas. *Jurnal Universal Penelitian Pendidikan*, 8 (11), 5312-5317. <https://doi.org/10.13189/ujer.2020.081132>