

Analisis Empiris Kualitas Butir Soal Penilaian Sumatif Mata Pelajaran Desain Komunikasi Visual Berbasis Evaluasi Pembelajaran di SMK Negeri 1 Mojokerto

Hasan Ismail^{*1}, Rufi'i²

^{1,2}Teknologi Pendidikan, Universitas PGRI Adi Buana Surabaya, Indonesia
Email: ¹250020071@student.unipasby.ac.id, ²rufii@unipasby.ac.id

Abstrak

Kualitas instrumen evaluasi memiliki peranan penting dalam menentukan tingkat akurasi pengukuran hasil belajar peserta didik. Namun, dalam praktiknya masih ditemukan penyusunan soal yang belum melalui proses analisis empiris sehingga validitas dan kualitas pengukuran belum optimal. Penelitian ini bertujuan untuk menganalisis kualitas butir soal Penilaian Sumatif Akhir Semester (PSAS) pada mata pelajaran Desain Komunikasi Visual (DKV) di SMK Negeri 1 Mojokerto. Penelitian menggunakan pendekatan kuantitatif deskriptif dengan teknik analisis butir soal terhadap 30 soal pilihan ganda. Analisis dilakukan menggunakan perangkat lunak Anates Versi 4.0.9 dengan mengkaji aspek validitas, tingkat kesukaran, daya pembeda, dan efektivitas distraktor. Hasil penelitian menunjukkan bahwa sebanyak 20 butir soal (66,7%) dinyatakan valid, sedangkan 10 butir soal (33,3%) belum memenuhi kriteria validitas. Tingkat kesukaran didominasi kategori sedang (60%), sedangkan daya pembeda mayoritas berada pada kategori cukup hingga baik (83,3%). Analisis distraktor menunjukkan bahwa sebagian besar pengecoh telah berfungsi secara efektif, meskipun beberapa opsi jawaban masih memerlukan perbaikan. Temuan penelitian mengindikasikan bahwa instrumen evaluasi secara umum telah memenuhi karakteristik tes yang baik, namun revisi diperlukan pada beberapa butir soal yang belum optimal. Hasil penelitian ini dapat dimanfaatkan sebagai dasar pengembangan bank soal dan peningkatan kompetensi guru dalam menyusun instrumen evaluasi yang valid, objektif, dan berkualitas pada pendidikan vokasi khususnya bidang Desain Komunikasi Visual.

Kata kunci: *analisis butir soal, daya pembeda, desain komunikasi visual, evaluasi pembelajaran, tingkat kesukaran, validitas.*

Empirical Analysis of Summative Assessment Item Quality in Visual Communication Design Subjects Based on Learning Evaluation in Vocational Schools Mojokerto

Abstract

The quality of assessment instruments plays an important role in determining the accuracy of measuring students' learning outcomes. However, in practice, test items are often developed without undergoing empirical item analysis, resulting in suboptimal validity and measurement quality. This study aims to analyze the quality of the End-of-Semester Summative Assessment (PSAS) items in the Visual Communication Design (VCD) subject at SMK Negeri 1 Mojokerto. The research employed a descriptive quantitative approach using item analysis techniques on 30 multiple-choice questions. The analysis was conducted using Anates Version 4.0.9 by examining four aspects: validity, difficulty level, discrimination index, and distractor effectiveness. The results revealed that 20 items (66.7%) were classified as valid, while 10 items (33.3%) did not meet the validity criteria. The difficulty level analysis showed that most items were in the moderate category (60%), while the discrimination index indicated that the majority of items were categorized as fair to good (83.3%). Distractor analysis demonstrated that most distractors functioned effectively, although several answer options still required revision. The findings indicate that the assessment instrument generally met the characteristics of a good test; however, improvements are needed for several items that have not yet achieved optimal quality. The results of this study can serve as a basis for developing an item bank and enhancing teachers' competencies in constructing valid, objective, and high-quality assessment instruments in vocational education, particularly in the field of Visual Communication Design.

Keywords: *item analysis, discrimination index, visual communication design, educational assessment, difficulty level, validity.*

1. PENDAHULUAN

Evaluasi pembelajaran merupakan bagian integral dalam sistem pendidikan karena berfungsi sebagai alat untuk mengukur ketercapaian kompetensi peserta didik secara objektif dan sistematis [1] [2]. Dalam konteks pendidikan vokasi, evaluasi tidak hanya berorientasi pada penguasaan konsep teoritis, tetapi juga pada kemampuan aplikatif dan pemecahan masalah yang relevan dengan kebutuhan industri [3] [4]. Oleh sebab itu, kualitas instrumen evaluasi menjadi faktor fundamental dalam menghasilkan data pembelajaran yang akurat dan dapat dipertanggungjawabkan secara akademik melalui pemenuhan aspek validitas dan reliabilitas instrumen [5] [6].

Salah satu bentuk evaluasi yang umum digunakan di Sekolah Menengah Kejuruan Negeri 1 Mojokerto adalah Penilaian Sumatif Akhir Semester (PSAS). Instrumen PSAS memiliki peranan strategis karena digunakan sebagai dasar penentuan capaian kompetensi peserta didik setelah mengikuti proses pembelajaran selama satu semester [7] [8]. Namun demikian, masih banyak instrumen tes yang disusun tanpa melalui proses analisis empiris terhadap kualitas butir soal. Kondisi tersebut berpotensi menghasilkan instrumen yang kurang valid, kurang mampu membedakan kemampuan peserta didik, serta memiliki tingkat kesukaran yang tidak proporsional [9]. Oleh karena itu, analisis butir soal diperlukan untuk memastikan bahwa instrumen memiliki validitas, daya pembeda, tingkat kesukaran, dan reliabilitas yang memadai sehingga mampu menghasilkan informasi yang akurat mengenai capaian belajar peserta didik [6].

Pada mata pelajaran Desain Komunikasi Visual (DKV), kualitas soal menjadi semakin penting karena karakteristik pembelajarannya menuntut kemampuan berpikir kreatif, analitis, visual, dan problem solving. Instrumen evaluasi yang baik harus mampu mengukur keterampilan tersebut secara tepat sehingga hasil evaluasi benar-benar mencerminkan kompetensi peserta didik [10] [11].

Analisis butir soal merupakan salah satu pendekatan yang dapat digunakan untuk mengidentifikasi kualitas instrumen evaluasi secara empiris. Melalui analisis butir soal, guru dapat mengetahui validitas, tingkat kesukaran, serta daya pembeda setiap soal sehingga dapat menentukan apakah suatu butir soal layak digunakan, direvisi, atau dibuang [1] [12]. Analisis ini juga menjadi bagian penting dalam pengembangan budaya evaluasi berbasis data di lingkungan pendidikan vokasi karena mampu meningkatkan akurasi pengukuran kompetensi peserta didik [13] [14].

Beberapa penelitian terdahulu umumnya berfokus pada analisis validitas, tingkat kesukaran, dan daya pembeda instrumen evaluasi pada mata pelajaran umum maupun pendidikan vokasi [15] [16] [17]. Namun demikian, kajian yang secara khusus menganalisis kualitas instrumen evaluasi pada mata pelajaran Desain Komunikasi Visual (DKV) di SMK masih relatif terbatas. Selain itu, sebagian besar penelitian terdahulu belum memasukkan analisis efektivitas distraktor sebagai bagian integral dalam evaluasi kualitas soal pilihan ganda. Padahal, efektivitas distraktor merupakan indikator penting yang menunjukkan kemampuan suatu butir soal dalam mengidentifikasi perbedaan kemampuan peserta didik secara lebih akurat [18] [19] [20].

Berdasarkan kondisi tersebut, penelitian ini memiliki kebaruan (novelty) pada analisis empiris kualitas butir soal PSAS mata pelajaran DKV yang mencakup empat indikator utama, yaitu validitas, tingkat kesukaran, daya pembeda, dan efektivitas distraktor. Hasil penelitian diharapkan dapat memberikan kontribusi empiris terhadap pengembangan instrumen evaluasi berbasis data pada pendidikan vokasi, khususnya pada bidang Desain Komunikasi Visual [21] [17].

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan metode analisis butir soal. Pendekatan ini digunakan untuk mendeskripsikan kualitas instrumen evaluasi berdasarkan data empiris hasil pengerjaan peserta didik [22] [23].

Objek penelitian berupa 30 butir soal pilihan ganda Penilaian Sumatif Akhir Semester (PSAS) mata pelajaran Desain Komunikasi Visual. Subjek penelitian adalah peserta didik kelas XII SMK Negeri 1 Mojokerto yang mengikuti evaluasi pembelajaran.

Teknik pengumpulan data dilakukan melalui dokumentasi lembar jawaban peserta didik. Data kemudian dianalisis untuk memperoleh informasi mengenai validitas, tingkat kesukaran, daya pembeda, dan efektivitas distraktor setiap butir soal [15] [24].

Analisis data dilakukan menggunakan perangkat lunak Anates Versi 4.0.9 yang banyak digunakan dalam analisis butir soal berbasis Teori Tes Klasik (*Classical Test Theory*) [25] [26].

Kriteria validitas ditentukan berdasarkan koefisien korelasi *product moment*, sedangkan tingkat kesukaran dan daya pembeda dianalisis berdasarkan indeks yang diperoleh dari hasil pengerjaan peserta didik [15]. Butir soal dinyatakan valid apabila memiliki koefisien korelasi $\geq 0,30$. Tingkat kesukaran diklasifikasikan menjadi kategori sukar, sedang, dan mudah. Sementara itu, daya pembeda dikategorikan menjadi jelek, cukup, dan baik.

Tahapan penelitian meliputi:

1. Pengumpulan data hasil tes peserta didik

2. Penginputan data ke aplikasi analisis butir soal
 3. Pengolahan data statistik butir soal
 4. Interpretasi hasil analisis
 5. Penarikan kesimpulan kualitas instrumen
- Efektivitas distraktor ditentukan berdasarkan kemampuan setiap pengecoh dalam menarik peserta didik berkemampuan rendah dan dipilih oleh proporsi peserta tes yang memadai [18] [19].

3. HASIL DAN PEMBAHASAN

3.1. Analisis Validitas Butir Soal

Validitas merupakan indikator utama dalam menentukan kemampuan soal untuk mengukur kompetensi yang seharusnya diukur. Hasil analisis menunjukkan bahwa dari 30 butir soal yang dianalisis, sebanyak 20 butir soal dinyatakan valid dan 10 butir soal belum memenuhi kriteria validitas.

Tabel 1. Hasil Validitas Butir Soal

Kategori	Jumlah	Persentase
Valid	20	66,7%
Tidak Valid	10	33,3%

Hasil tersebut menunjukkan bahwa sebagian besar instrumen telah memiliki kualitas pengukuran yang baik. Tingginya persentase soal valid mengindikasikan bahwa mayoritas butir soal mampu merepresentasikan kompetensi yang diukur pada mata pelajaran DKV.

Namun demikian, masih terdapat beberapa soal yang belum valid. Hal ini dapat disebabkan oleh ketidaksesuaian indikator soal, redaksi soal yang ambigu, atau distraktor yang kurang efektif. Butir soal yang belum valid memerlukan revisi agar kualitas instrumen menjadi lebih optimal.

3.2. Analisis Tingkat Kesukaran

Tingkat kesukaran menunjukkan proporsi peserta didik yang mampu menjawab soal dengan benar. Hasil analisis menunjukkan bahwa mayoritas soal berada pada kategori sedang.

Tabel 2. Distribusi Tingkat Kesukaran Soal

Kategori	Jumlah
Sukar	6
Sedang	18
Mudah	6

Dominasi soal kategori sedang menunjukkan bahwa instrumen memiliki tingkat kesukaran yang proporsional. Soal kategori sedang dinilai ideal karena mampu memberikan tantangan akademik tanpa menyebabkan peserta didik mengalami kesulitan yang berlebihan.

Keberadaan soal kategori mudah dan sukar tetap diperlukan sebagai variasi tingkat kognitif serta untuk mengakomodasi perbedaan kemampuan peserta didik.

3.3. Analisis Daya Pembeda

Daya pembeda merupakan kemampuan soal dalam membedakan peserta didik yang memiliki kemampuan tinggi dan rendah.

Tabel 3. Distribusi Daya Pembeda Soal

Kategori	Jumlah
Jelek	5
Cukup	12
Baik	13

Hasil analisis menunjukkan bahwa sebagian besar soal memiliki daya pembeda cukup hingga baik. Hal ini mengindikasikan bahwa instrumen mampu mengidentifikasi perbedaan kemampuan peserta didik secara efektif.

Soal dengan daya pembeda rendah perlu direvisi karena kurang mampu membedakan kemampuan peserta didik secara signifikan. Faktor penyebab rendahnya daya pembeda dapat berasal dari kualitas distraktor yang kurang berfungsi atau tingkat kesukaran yang tidak proporsional.

3.4. Analisis Efektivitas Distraktor

Tabel 4. Efektivitas Distraktor	
Kategori	Jumlah Soal
Efektif	22
Kurang Efektif	8

Analisis efektivitas distraktor menunjukkan bahwa sebagian besar butir soal telah memiliki pengecoh yang berfungsi dengan baik. Sebanyak 22 butir soal memiliki distraktor efektif, sedangkan 8 butir soal masih memiliki satu atau lebih opsi jawaban yang kurang menarik bagi peserta didik. Kondisi tersebut menunjukkan bahwa sebagian besar pengecoh telah mampu menjalankan fungsinya untuk mengidentifikasi peserta didik yang belum menguasai materi pembelajaran. Namun demikian, beberapa distraktor perlu direvisi agar mampu meningkatkan kualitas pengukuran dan daya pembeda soal.

3.5. Pembahasan

Secara umum, kualitas instrumen PSAS mata pelajaran DKV telah memenuhi karakteristik tes yang baik. Mayoritas soal memiliki validitas yang memadai, tingkat kesukaran sedang, dan daya pembeda cukup hingga baik. Temuan ini menunjukkan bahwa instrumen evaluasi telah mampu mengukur kemampuan peserta didik secara relatif objektif.

Dalam konteks pendidikan vokasi, instrumen evaluasi yang baik memiliki peranan penting dalam mendukung pembelajaran berbasis kompetensi. Pada mata pelajaran DKV, instrumen evaluasi tidak hanya mengukur kemampuan mengingat konsep, tetapi juga kemampuan memahami pesan visual, menganalisis desain, dan menerapkan prinsip estetika.

Analisis butir soal menjadi langkah strategis dalam meningkatkan kualitas evaluasi pembelajaran. Guru tidak hanya berperan sebagai penyusun soal, tetapi juga sebagai evaluator kualitas instrumen pembelajaran. Melalui analisis empiris, guru dapat melakukan perbaikan berkelanjutan terhadap instrumen evaluasi sehingga kualitas pembelajaran menjadi lebih terukur dan akuntabel.

Temuan penelitian ini juga memperkuat pentingnya pengembangan budaya evaluasi berbasis data di lingkungan SMK. Penggunaan analisis butir soal secara rutin dapat membantu sekolah menghasilkan instrumen evaluasi yang lebih profesional dan sesuai dengan standar penilaian pendidikan modern.

Temuan penelitian ini sejalan dengan hasil penelitian [9] yang menunjukkan bahwa dominasi soal berkategori sedang dan memiliki daya pembeda yang baik merupakan indikator kualitas instrumen yang memadai dalam konteks pendidikan vokasi. Hasil penelitian ini juga mendukung temuan [27] yang menyatakan bahwa analisis empiris melalui validitas, tingkat kesukaran, dan daya pembeda mampu mengidentifikasi kualitas instrumen secara lebih objektif dibandingkan hanya melalui telaah teoritis.

Dari aspek validitas, persentase soal valid sebesar 66,7% menunjukkan bahwa sebagian besar instrumen telah mampu mengukur kompetensi yang ditargetkan [15]. Namun, persentase tersebut masih lebih rendah dibandingkan beberapa penelitian sebelumnya yang melaporkan tingkat validitas instrumen di atas 75% [17], [15], [28]. Perbedaan ini mengindikasikan bahwa penyusunan soal pada mata pelajaran Desain Komunikasi Visual memiliki tantangan tersendiri karena harus mampu mengukur kemampuan visual, analitis, dan kreatif secara bersamaan [10], [29].

4. KESIMPULAN

Penelitian ini menunjukkan bahwa instrumen Penilaian Sumatif Akhir Semester mata pelajaran Desain Komunikasi Visual di SMK Negeri 1 Mojokerto secara umum telah memenuhi karakteristik instrumen evaluasi yang baik. Kualitas butir soal ditunjukkan oleh dominasi soal yang valid, tingkat kesukaran yang proporsional, daya pembeda yang memadai, serta sebagian besar distraktor yang berfungsi efektif. Hasil tersebut mengindikasikan bahwa instrumen telah mampu mendukung proses pengukuran hasil belajar secara objektif dan akurat [15], [16]. Meskipun demikian, masih ditemukan beberapa butir soal yang memerlukan perbaikan terutama pada aspek validitas dan efektivitas distraktor. Oleh karena itu, guru disarankan untuk melakukan analisis butir soal secara berkala sebelum instrumen digunakan kembali guna menjaga kualitas pengukuran dan memperbaiki kelemahan butir soal yang teridentifikasi [18], [20].

Penelitian selanjutnya dapat mengembangkan analisis menggunakan pendekatan *Item Response Theory* (IRT) atau melibatkan jumlah sampel yang lebih besar untuk memperoleh informasi kualitas instrumen yang lebih mendalam [30], [31].

DAFTAR PUSTAKA

- [1] D. Mardapi, *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Yogyakarta: Parama Publishing, 2020.
- [2] A. Sudijono, *Pengantar Evaluasi Pendidikan*. Jakarta: RajaGrafindo Persada, 2022.
- [3] E. Winther and F. Achtenhagen, "Measurement of Vocational Competencies—A Contribution to an International Large-Scale Assessment on Vocational Education and Training," *Empirical Research in Vocational Education and Training*, 2009.
- [4] M. Nurtanto, Z. Arifin, and H. Sofyan, "Development of Model for Professional Competency Assessment (PCA) in Vocational Education," *Journal of Technical Education and Training*, 2020.
- [5] H. Retnawati, *Analisis Kuantitatif Instrumen Penelitian*. Yogyakarta: Parama Publishing, 2021.
- [6] N. Kholis, B. Kartowagiran, and D. Mardapi, "Development and Validation of an Instrument to Measure a Performance of Vocational High School," *European Journal of Educational Research*, 2020.
- [7] Z. Arifin, *Evaluasi Pembelajaran: Prinsip, Teknik, dan Prosedur*. Bandung: Remaja Rosdakarya, 2021.
- [8] D. Mardapi, *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Yogyakarta: Parama Publishing, 2020.
- [9] A. Z. Fathin and H. Retnawati, "How Good Are the Characteristics of the Mathematics National Standardized School Examination Items in Vocational School?," *J. Phys. Conf. Ser.*, vol. 1581, 2020.
- [10] L. Greenstein, *Assessing 21st Century Skills: A Guide to Evaluating Mastery and Authentic Learning*. Thousand Oaks, CA: Corwin Press, 2012.
- [11] P. Sukkeewan, N. Songkram, and J. Nasongkhla, "Development and Validation of a Reliable and Valid Assessment Tool for Measuring Innovative Thinking in Vocational Students," *Journal of Educational Research and Practice*, 2024.
- [12] H. Retnawati, *Analisis Kuantitatif Instrumen Penelitian*. Yogyakarta: Parama Publishing, 2021.
- [13] N. Kholis, B. Kartowagiran, and D. Mardapi, "Development and Validation of an Instrument to Measure a Performance of Vocational High School," *European Journal of Educational Research*, vol. 9, no. 3, pp. 955–966, 2020.
- [14] H. Maksum, D. Yuvenda, and W. Purwanto, "Teaching Factory Based on Troubleshooting (TEFA-T) Model to Improve Problem-Solving and Critical Thinking Skills in Vocational Learning," *Journal of Turkish Science Education*, 2022.
- [15] D. Mardapi, *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Yogyakarta: Parama Publishing, 2020.
- [16] H. Retnawati, *Analisis Kuantitatif Instrumen Penelitian*. Yogyakarta: Parama Publishing, 2021.
- [17] N. Kholis, B. Kartowagiran, and D. Mardapi, "Development and Validation of an Instrument to Measure a Performance of Vocational High School," *European Journal of Educational Research*, vol. 9, no. 3, pp. 955–966, 2020.
- [18] M. J. Gierl, O. Bulut, and Q. Guo, "Developing, Analyzing, and Using Distractors for Multiple-Choice Tests in Education: A Comprehensive Review," *Rev. Educ. Res.*, vol. 87, no. 6, pp. 1082–1116, 2017, doi: 10.3102/0034654317726529.
- [19] M. Sajjad, S. Iltaf, and R. A. Khan, "Nonfunctional Distractor Analysis: An Indicator for Quality of Multiple Choice Questions," *Pak. J. Med. Sci.*, vol. 36, no. 5, pp. 982–986, 2020.
- [20] A. A. Rezigalla, A. M. E. S. A. Eleragi, and A. B. Elhusssein, "Item Analysis: The Impact of Distractor Efficiency on the Difficulty Index and Discrimination Power of Multiple-Choice Items," *BMC Med. Educ.*, vol. 24, 2024.
- [21] A. Friatma and A. Anhar, "Analysis of Validity, Reliability, Discrimination, Difficulty and Distraction Effectiveness in Learning Assessment," *J. Phys. Conf. Ser.*, vol. 1387, 2019.
- [22] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. 2022.
- [23] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 2018.
- [24] H. Retnawati, *Analisis Kuantitatif Instrumen Penelitian*. 2021.

- [25] H. Sutrisno, "An Analysis of the Mathematics School Examination Test Quality," *Jurnal Riset Pendidikan Matematika*, vol. 3, no. 2, pp. 195–206, 2016.
- [26] A. T. Panjaitan and H. Retnawati, "The Characteristics of Final Mathematics Test Items Based on Classical Test Theory," *J. Phys. Conf. Ser.*, vol. 1397, 2019, doi: 10.1088/1742-6596/1397/1/012095.
- [27] P. I. Wiyasa, I. Laksana, and N. L. K. M. Indrawati, "Evaluating Quality of Teacher-Developed English Test in Vocational High School: Content Validity and Item Analysis," *Education Quarterly Reviews*, vol. 2, no. 4, 2019.
- [28] M. Suhaini, A. Ahmad, and N. M. Bohari, "Assessments on Vocational Knowledge and Skills: A Content Validity Analysis," *European Journal of Educational Research*, 2021.
- [29] T. Simon, I. Biro, and A. Karpati, "Developmental Assessment of Visual Communication Skills in Primary Education," *J. Intell.*, vol. 10, no. 3, 2022.
- [30] T. M. Haladyna, "Item Analysis for Selected-Response Test Items," in *Handbook of Test Development*, 3rd ed., Routledge, 2015.
- [31] R. K. Hambleton, H. Swaminathan, and H. J. Rogers, *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Publications, 1991.