

VALIDITAS DAN RELIABILITAS INSTRUMEN TES DIAGNOSTIK *FOUR-TIER MULTIPLE CHOICE* (4TMC) PADA KONSEP LAJU REAKSI

VALIDITY AND RELIABILITY OF *FOUR-TIER MULTIPLE CHOICE* (4TMC) DIAGNOSTIC TEST INSTRUMENT ON REACTION RATE CONCEPT

Ilham Pradana Putra Harahap dan *Dian Novita

Jurusan Kimia FMIPA Universitas Negeri Surabaya

e-mail: diannovita@unesa.ac.id

Abstrak

Miskonsepsi didefinisikan sebagai pandangan atau ide yang salah mengenai sebuah konsep. Pandangan tersebut berbeda dengan konsep yang telah diyakini oleh para ahli kebenarannya. Miskonsepsi yang dimiliki oleh peserta didik tidak dapat dibiarkan begitu saja. Miskonsepsi harus dapat dideteksi sedini mungkin agar tidak memengaruhi pemahaman peserta didik terhadap konsep-konsep berikutnya yang saling memiliki keterkaitan antar konsep. Tes diagnostik dapat digunakan untuk mendeteksi miskonsepsi. Penelitian ini bertujuan untuk menentukan kelayakan instrumen tes diagnostik *four-tier multiple choice* berdasarkan aspek validitas dan reliabilitas. Validitas instrumen tes diagnostik *four-tier multiple choice* dinilai oleh tiga validator berdasarkan aspek isi, konstruk dan bahasa. Reliabilitas tes diagnostik *four-tier multiple choice* dihitung menggunakan koefisien *Alpha Cronbach*. Berdasarkan hasil penelitian, dapat disimpulkan bahwa (1) validitas instrumen ditinjau dari aspek isi, konstruk dan bahasa berturut-turut sebesar 88,35%, 90,03% dan 95,6% dengan kriteria sangat valid dan (2) reliabilitas instrumen yang dihitung sebesar 0,714 yang artinya $r_{hitung} > r_{tabel}$. Instrumen tes diagnostik yang dikembangkan dinyatakan layak untuk digunakan.

Kata Kunci: Validitas, Reliabilitas, *Four-Tier Multiple Choice*

Abstract

Misconception is defined as a wrong view or idea of a concept. This view is different from the concept that has been believed by experts to the truth. Misconceptions held by students cannot be left alone. Misconceptions must be detected as early as possible so as not to affect learners' understanding of subsequent concepts that have interrelationships between concepts. Diagnostic tests can be used to detect misconceptions. The aims of this study were to produce a four-tier multiple choice diagnostic test instrument that is worth based on aspects of validity and reliability. The validity of the four-tier multiple diagnostic test instrument was assessed by three validators in terms of content, construct and language. The Reliability of the four-tier multiple diagnostic test was calculated using the Alpha Cronbach coefficient. Based on the results of the study, it can be concluded that (1) the validity of the instrument in terms of content, construct and language are respectively 88.35%, 90.03% and 95.6% with very valid criteria, and (2) the reliability of the instrument calculated is 0.714, which means $r_{count} > r_{table}$. The developed diagnostic test instrument can be declared worth for use.

Keywords: Validity, Reliability, *Four-Tier Multiple Choice*

PENDAHULUAN

Ilmu kimia merupakan ilmu yang mempelajari berbagai hal seperti komposisi, struktur, sifat, perubahan serta energi yang menyertai dari suatu materi. Ilmu ini dikembangkan menggunakan metode ilmiah dan didasari oleh sikap ilmiah. Pemahaman konsep dalam pembelajaran kimia sangatlah penting karena terdapat keterkaitan antar konsepnya. Konsep dalam ilmu kimia harus didapatkan oleh peserta didik sebagai landasan yang kuat untuk dapat melakukan pembelajaran aplikatif berikutnya.

Miskonsepsi didefinisikan sebagai pandangan atau ide yang salah mengenai sebuah konsep. Pandangan tersebut berbeda dengan konsep yang telah diyakini oleh para ahli kebenarannya. Miskonsepsi yang terjadi pada peserta didik biasanya bersifat resisten dan persisten [1]. Oleh karena itu miskonsepsi harus dideteksi sejak awal agar pemahaman konsep-konsep yang akan dipelajari peserta didik selanjutnya tidak terganggu [2].

Hasil penelitian sebelumnya mengenai miskonsepsi dalam mata pelajaran kimia yaitu penelitian yang dilakukan oleh [3] yang

menghasilkan penelitian temuan bahwa peserta didik kelas XI IPA 1 SMAN 1 Lhokseumawe terdeteksi mengalami miskonsepsi pada beberapa faktor yang memengaruhi laju reaksi. Peserta didik sebanyak 13,16% terdeteksi mengalami miskonsepsi pada faktor luas permukaan bidang sentuh terhadap laju reaksi. Sedangkan pada pengaruh faktor suhu dan katalis jumlah peserta didik mengalami miskonsepsi dengan persentase yang sama yaitu 57,89%. Namun pengaruh faktor konsentrasi terhadap laju reaksi tidak ada peserta didik yang mengalami miskonsepsi.

Tes diagnostik dapat digunakan sebagai suatu metode untuk mengetahui jika peserta didik mengalami miskonsepsi [4]. Kelemahan dan kekuatan yang dimiliki oleh peserta didik dapat diketahui melalui hasil tes diagnostik yang telah dilakukan. Kelemahan tersebut akan diberikan perlakuan yang tepat dan tindak lanjut yang sesuai. Guru dapat menggunakan tes diagnostik di awal maupun di akhir pembelajaran guna mendeteksi miskonsepsi yang dialami peserta didik pada materi yang dipelajari [5].

Penyusunan instrumen tes diagnostik *four-tier multiple choice* ini didasarkan pada empat tingkatan soal. Tingkat pertama adalah soal dalam bentuk soal pilihan ganda. Soal pada tingkat ke dua adalah tingkat keyakinan peserta didik dalam memilih jawaban. Soal pada tingkat ke tiga adalah alasan peserta didik dalam menjawab pertanyaan. Sedangkan pada tingkat ke empat merupakan soal mengenai tingkat keyakinan peserta didik dalam memilih alasan.

Kelayakan instrumen tes diagnostik ditinjau berdasarkan 2 aspek, yaitu validitas dan reliabilitas. Validitas yang akan diukur dalam instrumen tes ini meliputi validitas isi, validitas konstruk dan validitas bahasa. Validasi merupakan kegiatan yang dilakukan oleh ahli untuk menentukan kelayakan suatu instrumen dalam mengukur apa yang akan diukur [6], dimana yang dimaksud dalam penelitian ini yakni miskonsepsi peserta didik. Reliabilitas yakni tingkat keajegan sebuah soal dalam menilai apa yang akan dinilai. Ajeg yang dimaksud yaitu tes menghasilkan hasil yang mengikuti perubahan secara ajeg pula. Sebuah instrumen tes dikatakan

baik apabila instrumen tersebut dapat dengan ajeg memberikan data sesuai dengan kenyataan.

METODE

Penelitian ini memiliki tujuan untuk menentukan kelayakan instrumen tes diagnostik *four-tier multiple choice* berdasarkan aspek validitas dan reliabilitas.

Jenis penelitian yang digunakan dalam adalah Penelitian dan Pengembangan (*Research and Development / R&D*) modifikasi [7]. Langkah penelitian yang dilakukan sampai tahap 3D yang meliputi *define, design* dan *develop*. Kegiatan yang dilakukan pada tahap *define* yaitu menganalisis kebutuhan serta kajian literatur. Tahap kedua yaitu tahap *design*. Kegiatan yang dilakukan adalah penentuan materi, penulisan kisi-kisi soal tes, penentuan bentuk tes dan penulisan soal tes. Tahap *develop* yang dilakukan yaitu penelaahan soal tes, revisi soal tes, validasi soal tes serta pengujian instrumen secara terbatas. Subjek penelitian yang digunakan saat uji coba terbatas yaitu sebanyak 33 peserta didik kelas XI IPA 1 SMAN 1 Driyorejo yang merupakan kelas heterogen dan telah menerima materi laju reaksi sebelumnya.

Instrumen tes diagnostik yang telah disusun harus dianalisis validitasnya. Data validitas instrumen tes diagnostik yang diperoleh dari lembar validasi yang dinilai oleh 3 validator yang terdiri dari dua dosen Pendidikan Kimia Unesa selaku dosen penguji dan seorang guru kimia di SMA Negeri 1 Driyorejo Gresik. Hasil validasi tersebut selanjutnya dianalisis secara kuantitatif menggunakan perhitungan persentase menggunakan skala Likert yang dapat diamati pada Tabel 1

Tabel 1. Skala Likert

Penilaian	Nilai Skala
Sangat Baik	5
Baik	4
Sedang	3
Buruk	2
Sangat Buruk	1

[8]

Nilai skala yang didapatkan berdasarkan Skala Likert pada Tabel 1 kemudian dihitung dengan menggunakan rumus untuk menentukan

persentase hasil validasi dari masing-masing kriteria adalah sebagai berikut:

$$P(\%) = \frac{\sum \text{skor total validator untuk masing - masing kriteria}}{\text{skor maksimal kriteria} \times \text{jumlah validator} \times 100\%}$$

Selanjutnya untuk mengetahui validitas instrumen, hasil dalam bentuk persentase selanjutnya dikembangkan dan diinterpretasikan berdasarkan kriteria skor yang tertuang pada Tabel 2.

Tabel 2. Interpretasi Skor

Persentase (%)	Kriteria
0-20	Sangat Kurang
21-40	Kurang
41-60	Cukup
61-80	Valid
81-100	Sangat Valid

[8]

Berdasarkan kriteria pada Tabel 2, instrumen tes diagnostik dikatakan valid apabila persentasenya $\geq 61\%$.

Reliabilitas instrumen diperoleh melalui hasil jawaban peserta didik saat uji coba terbatas. Setiap peserta didik yang menjawab dengan benar pada soal tingkat 1 dan 3 serta menjawab yakin pada tingkat 2 dan 4 (kategori paham konsep) akan diberikan skor 1 (satu), sedangkan selain kombinasi jawaban tersebut diberikan nilai 0 (nol). Reliabilitas tes dapat dihitung menggunakan persamaan *Alpha Cronbach* yang dalam [9] adalah sebagai berikut,

$$r = \frac{k}{k-1} \left(1 - \frac{\sum Si^2}{St^2} \right)$$

$$Si^2 = \frac{\sum Xi^2 \frac{(\sum Xi)^2}{n}}{n}$$

$$St^2 = \frac{\sum X^2 \frac{(\sum X)^2}{n}}{n}$$

Instrumen dikatakan memiliki reliabilitas yang baik apabila $r_{hitung} > r_{tabel}$ [10].

HASIL DAN PEMBAHASAN

Validitas Tes

Validitas yang diukur pada instrumen tes ini meliputi validitas isi, validitas konstruk serta validitas bahasa. Validitas tes dinilai tiap butir soal agar dapat mengetahui secara spesifik terhadap kelebihan dan kekurangan soal tes.

Validitas isi menilai kesesuaian tes terhadap materi yang diujikan. Validitas konstruk menilai kesesuaian tes terhadap komponen-komponen yang harus terdapat pada tes. Validitas bahasa menilai kesesuaian bahasa yang digunakan terhadap kemampuan kognitif dan perkembangan psikologi peserta didik yang akan menjadi subjek penelitian.

Validitas Isi

Validitas isi meliputi: (1) kesesuaian antar butir soal dengan submateri pada faktor-faktor yang memengaruhi laju reaksi, (2) konsistensi antara setiap butir soal dengan indikator soal, (3) konsistensi setiap soal dengan urutan materi dan (4) kejelasan batasan antara setiap pertanyaan, dengan jawaban, dan penjelasan alasan yang diharapkan. Tabel 3 berikut merupakan perhitungan hasil penilaian validitas isi instrumen tes diagnostik *four-tier multiple choice* (4TMC).

Tabel 3. Perhitungan Hasil Penilaian Validitas Isi Instrumen Tes Diagnostik *Four-Tier Multiple Choice* (4TMC).

No.	Aspek yang Divalidasi	Rata-rata	% Validitas	Kriteria Validitas
1.	Kesesuaian antar butir soal dengan submateri pada faktor-faktor yang memengaruhi laju reaksi	4,67	88,3 5	Sangat Valid
2.	Konsistensi antara setiap butir soal dengan indikator soal	4,67		
3.	Konsistensi setiap soal dengan urutan materi	4,33		
4.	Kejelasan batasan antara setiap pertanyaan, dengan jawaban, dan penjelasan alasan yang	4		

No.	Aspek yang Divalidasi	Rata-rata	% Validitas	Kriteria Validitas
-----	-----------------------	-----------	-------------	--------------------

diharapkan

Hasil penilaian validitas dari Tabel 3 dapat dilihat bahwa validitas isi instrumen tes diagnostik ini sebesar 88,35% dan diinterpretasikan sesuai pedoman pada Tabel 2 menghasilkan kriteria sangat valid [6].

Validitas Konstruk

Validitas konstruk meliputi: (1) kejelasan petunjuk tes diagnostik, (2) kesesuaian butir soal dengan indikator soal serta kompetensi dasar, (3) setiap butir soal dapat mengidentifikasi miskonsepsi pada peserta didik, (4) pilihan alasan yang disajikan dapat mengungkapkan penyebab miskonsepsi yang bersumber dari peserta didik, (5) pilihan alasan yang merupakan pengecoh bersifat rasional dan homogen dengan jawaban tingkat pertama dan (6) kesesuaian tabel, grafik, gambar, dan sejenisnya dengan masalah yang disajikan. Tabel 4 berikut merupakan perhitungan hasil penilaian validitas konstruk instrumen tes diagnostik *four-tier multiple choice* (4TMC).

Tabel 4. Perhitungan Hasil Penilaian Validitas Konstruk Instrumen Tes Diagnostik *Four-Tier Multiple Choice* (4TMC).

No.	Aspek yang Divalidasi	Rata-rata	% Validitas	Kriteria Validitas
1.	Kejelasan petunjuk tes diagnostik	4,67	90,03	Sangat Valid
2.	Kesesuaian antara butir soal dengan indikator soal dan kompetensi dasar	4,33		
3.	Setiap butir tes diagnostik dapat mengidentifikasi miskonsepsi pada peserta didik	4		
4.	Pilihan alasan yang merupakan pengecoh bersifat rasional	4,67		

No.	Aspek yang Divalidasi	Rata-rata	% Validitas	Kriteria Validitas
-----	-----------------------	-----------	-------------	--------------------

dan homogen dengan jawaban tingkat pertama

5. Pengecoh pada pilihan alasan bersifat rasional dan homogen dengan jawaban tingkat pertama

6. Kesesuaian tabel, grafik, gambar, dan sejenisnya dengan masalah yang disajikan

Berdasarkan Tabel 4, validitas konstruk instrumen tes diagnostik memiliki kriteria yang sangat valid dengan persentase sebesar 90,03%.

Validitas Bahasa

Validitas bahasa meliputi: (1) kalimat soal menggunakan Bahasa Indonesia yang baik dan benar, (2) pertanyaan serta pernyataan yang digunakan tidak menimbulkan penafsiran ganda dan (3) pertanyaan pada setiap butir soal tes dinyatakan secara jelas serta komunikatif. Tabel 5 berikut merupakan perhitungan hasil penilaian validitas bahasa instrumen tes diagnostik *four-tier multiple choice* (4TMC).

Tabel 5. Perhitungan Hasil Penilaian Validitas Bahasa Instrumen Tes Diagnostik *Four-Tier Multiple Choice* (4TMC).

No.	Aspek yang Divalidasi	Rata-rata	% Validitas	Kriteria Validitas
1.	Kalimat soal menggunakan Bahasa Indonesia yang baik dan benar	5	95,6	Sangat Valid
2.	Pertanyaan serta pernyataan yang digunakan tidak menimbulkan penafsiran	4,67		

No.	Aspek yang Divalidasi	Rata-rata	% Validitas	Kriteria Validitas
3.	Pertanyaan pada butir-butir tes dinyatakan secara jelas serta komunikatif	4,67		

Berdasarkan Tabel 5 dapat dilihat bahwa hasil penilaian validitas konstruk yaitu sebesar 95,6% dan diinterpretasikan sesuai pedoman pada Tabel 2 menghasilkan kriteria sangat valid [8].

Berdasarkan hasil penilaian validitas yang ditinjau berdasarkan aspek isi, konstruk dan bahasa, maka dapat disederhanakan seperti pada Tabel 6 berikut.

Tabel 6. Hasil Validitas Tes

Aspek	Persentase(%)	Kriteria
Isi	88,35	Sangat valid
Konstruk	90,03	Sangat valid
Bahasa	95,6	Sangat valid

Berdasarkan Tabel 6 dapat diketahui bahwa instrumen diagnostik *four-tier multiple choice* (4TMC) pada Konsep Laju Reaksi yang sudah dikembangkan dan diuji ternasuk ke dalam kriteria sangat valid dengan persentase hasil validitas isi, validitas konstruk dan validitas bahasa berdasarkan Tabel 6 berturut-turut adalah 88,35%; 90,03% dan 95,6%.

Reliabilitas Tes

Berdasarkan hasil perhitungan reliabilitas menggunakan persamaan *Alpha Cronbach*, didapatkan hasil nilai reliabilitas instrumen (r_{hitung}) sebesar 0,714 yang dapat dilihat dalam Tabel 7 di bawah ini.

Tabel 7. Hasil Perhitungan Reliabilitas Tes Menggunakan SPSS

Item-Total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
soal1	14.64	7.489	.484	.693
soal2	14.67	7.604	.311	.702
soal3	14.67	7.229	.557	.684

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
soal4	14.97	8.155	-.075	.747
soal5	14.64	7.489	.484	.693
soal6	14.91	7.898	.023	.734
soal7	14.61	8.121	.038	.717
soal8	14.73	8.392	-.159	.741
soal9	14.85	6.883	.465	.683
soal10	14.64	7.676	.339	.702
soal11	14.85	6.570	.610	.665
soal12	14.82	7.028	.421	.689
soal13	14.64	7.301	.634	.683
soal14	14.85	7.883	.040	.730
soal15	15.55	8.068	.092	.715
soal16	14.73	6.892	.609	.673
soal17	14.76	7.439	.279	.704
soal18	14.88	7.360	.242	.709
soal19	15.00	6.375	.617	.661
soal20	14.58	8.189	.000	.716

Reliability Statistics

Cronbach's Alpha	N of Items
.714	20

Adapun nilai r_{tabel} yang digunakan untuk $n=33$ dengan tingkat kepercayaan 95% yaitu sebesar 0,344. Dengan hasil demikian, hasil $r_{hitung} > r_{tabel}$ sehingga instrumen tes diagnostik *four-tier multiple choice* (4TMC) di konsep Laju Reaksi yang sudah dikembangkan serta diuji coba dapat dinyatakan reliabel [10].

Berdasarkan hasil dari aspek validitas dan reliabilitas dapat diketahui bahwa secara keseluruhan instrumen tes diagnostik *four-tier multiple choice* (4TMC) yang telah dikembangkan dapat dikatakan layak dan dapat digunakan untuk mengidentifikasi miskonsepsi peserta didik pada submateri faktor-faktor yang memengaruhi laju reaksi. Hal tersebut sesuai pernyataan [11] bahwa aspek validitas dan reliabilitas merupakan dua parameter yang dapat merepresentasikan kualitas dari instrumen secara keseluruhan.

PENUTUP

Simpulan

Instrumen tes diagnostik *Four-Tier Multiple Choice* (4TMC) pada konsep faktor-faktor yang memengaruhi Laju Reaksi yang sudah dikembangkan dapat dinyatakan telah layak dengan rincian seperti di bawah ini.

1. Validitas isi, konstruk dan bahasa yang diperoleh dari hasil penilaian tiga validator berturut-turut adalah 88,35%; 90,03% dan 95,6% yang tergolong kriteria sangat valid pada setiap aspek.
2. Instrumen yang dikembangkan dinyatakan reliabel dengan hasil $r_{hitung} > r_{tabel}$ dengan rincian nilai koefisien reliabilitas pada instrumen (r_{hitung}) sebesar 0,714 dan nilai r_{tabel} yang digunakan yaitu 0,344

Saran

1. Instrumen 4TMC yang telah dikembangkan ini dapat diteliti lebih lanjut sebagai upaya untuk mereduksi adanya miskonsepsi pada konsep laju reaksi.
2. Pengembangan dengan IT terhadap instrumen 4TMC ini agar lebih mudah digunakan oleh guru maupun peserta didik.

DAFTAR PUSTAKA

1. Ibrahim, M. 2012. *Seri Pembelajaran Inovatif: Konsep, Miskonsepsi dan Cara Pembelajarannya*. Surabaya: University Press.
2. Treagust, D. 2006. "Diagnostic Assessment in Science as a Means to Improving Teaching, Learning and Retention." In UniServe Science Assessment Symposium, Sep 28, 2006, Sydney, NSW: UniServe Science, The University of Sydney.
3. Nazar, M., Sulastri, Winarni, S., & Fitriani, R. 2013. Identifikasi Miskonsepsi Peserta Didik SMA pada Konsep Faktor-Faktor yang Memengaruhi Laju Reaksi. *Jurnal Biologi Edukasi*, 2(3), 49-53.
4. Depdiknas. 2007. *Tes Diagnostik, Direktorat Pembinaan Sekolah Menengah Pertama*. Jakarta: Direktorat Jenderal Manajemen Pendidikan Dasar dan Menengah.
5. Lin, S. 2004. Development and Application of a Two Tier Diagnostic Test for High School Students' Understanding of Sowering Plant Growth and Development. *International Journal of Science and Mathematics* , 2(2), 175-199.
6. Sugiyono. 2008. *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
7. Thiagarajan, S., Semmel, D. S., & Semmel, M. I. 1974. *Instructional Development for Training Teachers of Exceptional Children: A Sourcebook*. Indiana: Indiana University.
8. Riduwan. 2015. *Skala Pengukuran Variabel-variabel Penelitian*. Bandung: Alfabeta.
9. Siregar, S. 2013. *Metode Penelitian Kuantitatif*. Jakarta: Kencana.
10. Arikunto, S. 2005. *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
11. Kimberlin, C. L., & Winterstein, A. G. 2008. Validity and Reliability of Measurement Instruments Used in Research. *Am J Health Syst Pharm* , 65(23) 2276-2284.