

Rekonstruksi *Diagnostic Test Four Tier* Berbasis HOTS (*High Order Thinking Skill*) Pada Materi Fiqih Kelas X di MA Negeri 3 Jombang

Fatati Nuriyana Mahmudah¹, Ino Angga Putra^{2*}, Muhammad Aliyul Wafa³, Andi Franoto⁴

^{1,3,4} Pendidikan Agama Islam, Universitas KH. A. Wahab Hasbullah

² Pendidikan Fisika, Universitas KH. A. Wahab Hasbullah

*Email: inoanggaputra@unwaha.ac.id

ABSTRACT

This research was conducted to develop an assessment instrument diagnostic test four tier. Diagnostic test four tier is one of the instruments used to diagnose student's level of conception on a concept. The purpose of this study was to design a four-tier diagnostic test assessment instrument based on high order thinking skills (HOTS) and also to determine the feasibility of understanding student's concept that were valid and reliable. This research is a development research that uses a 4-D model, namely define, design, develop, and disseminate. The data analysis technique uses validation, validity, reliability and difficulty level and using SPSS 21 calculations. The results of the overall validation obtained that the average score of the four-tier diagnostic test instrument was 3.04 and valid or suitable for use. The results of the instrument validity test obtained 17 valid questions and the results of the instrument reliability test obtained a value of 0.901 in the very high category. Therefore, the four-tier diagnostic test assessment instrument based on high order thinking skills (HOTS) on fiqh material for class X MA Negeri 3 Jombang deserves to be tested in the learning process.

Keywords: *Assessment Instrument, Four Tier Diagnostic Test Four Tier, High Order Thinking Skills, Concept Understanding*

ABSTRAK

Penelitian ini dilakukan untuk mengembangkan sebuah instrumen penilaian diagnostic test four tier. Diagnostic test four tier merupakan salah satu instrumen yang digunakan untuk mendiagnosis level konsepsi siswa pada suatu konsep. Tujuan dari penelitian ini adalah untuk mendesain instrumen penilaian diagnostic test four tier berbasis high order thinking skills (HOTS) dan juga untuk mengetahui kelayakan pemahaman konsep siswa yang valid dan reliabel. Penelitian ini adalah penelitian pengembangan yang menggunakan model 4-D yaitu define (pendefinisian), design (perancangan), develop (pengembangan), dan disseminate (penyebarluasan). Teknik analisis data menggunakan validasi, validitas, reliabilitas dan tingkat kesukaran soal serta menggunakan perhitungan SPSS 21. Hasil validasi secara keseluruhan diperoleh skor rata-rata instrumen diagnostic test four tier adalah 3.04 dan valid atau layak digunakan. Hasil uji validitas instrumen diperoleh 17 butir soal yang valid dan hasil uji reliabilitas instrumen diperoleh nilai 0.901 pada kategori sangat tinggi. Oleh karena itu, instrumen penilaian diagnostic test four tier berbasis high order thinking skills (HOTS) pada materi fiqih kelas X di MA Negeri 3 Jombang layak untuk diuji cobakan pada proses pembelajaran.

Kata Kunci: *Instrumen Penilaian, Diagnostic Test Four Tier, High Order Thinking Skills (HOTS), Pemahaman Konsep*

PENDAHULUAN

Proses kegiatan belajar dan mengajar cenderung mengarah pada hasil pembelajaran dimana erat kaitannya dengan pemahaman konsep. Pemahaman konsep dapat diartikan sebagai kemampuan siswa dalam memahami makna secara ilmiah, baik secara konsep maupun teori serta penerapannya dalam

kehidupan sehari-hari. Hubungan antara konsep-konsep dalam suatu materi pelajaran dapat diwujudkan dalam bentuk rumus-rumus untuk memecahkan masalah, grafik, bagan, poster, tabel, dan bentuk hubungan lain. Penguasaan konsep juga merupakan bagian dari hasil belajar pada komponen pembelajaran karena konsep, prinsip, dan struktur pengetahuan dan pemecahan masalah merupakan bagian hasil belajar pada ranah kognitif (Putra et al., 2018).

Memahami konsep merupakan kemampuan seseorang untuk menghasilkan ide-ide abstrak yang memungkinkan objek untuk dikategorikan/digolongkan. Setiap konsep bersifat berjenjang, hal ini dapat dilihat pada contoh konsep fungsi bijektif yang dikembangkan dari konsep relasi dan sebagainya (Setiawan, 2008). Konsep disusun sedemikian rupa sehingga konsep sebelumnya digunakan untuk mempelajari konsep selanjutnya. Oleh karena itu, agar siswa memahami situasi, pembelajaran perlu memahami makna dan filosofinya, maksud dan implikasinya, serta penerapannya (Agustin et al., 2018). Keberhasilan siswa dalam memahami konsep yang lebih kompleks didasarkan pada pemahaman konsep dasar yang mereka peroleh (Ulva et al., 2016).

Pemahaman konsep tersebut belum dapat dipastikan dari ketuntasan nilai yang diperoleh peserta didik. Hal tersebut dikarenakan evaluasi pembelajaran yang dilakukan belum semuanya dapat mendeteksi tingkat pemahaman konsep. Evaluasi Penilaian biasanya dilakukan pada ulangan harian, dan kesalahan yang dilakukan siswa saat menjawab soal biasanya hanya dinilai sebagai kesalahan perhitungan, kurang ketelitian, atau kurang belajar. Namun, guru tidak menemukan kemungkinan bahwa siswa akan salah memahami konsep materi. Soal pilihan ganda mudah digunakan, jadi disarankan agar anda menggunakan tes pilihan ganda untuk mempelajari penilaian, tetapi evaluator akan kesulitan dalam menentukan jawaban peserta didik tersebut menggambarkan tingkat kemampuannya ataukah jawaban tersebut hanya tebakan saja (Siswaningsih, 2014).

Pemahaman konsep sendiri tidak jauh dari adanya miskonsepsi atau kesalahpahaman. Menurut (Rohmah et al., 2019) bahwa miskonsepsi atau kesalahpahaman ini muncul dalam pendekatan konstruktivis, proses pembelajaran yang menggunakan pengetahuan yang dibentuk atas dasar kontak lingkungan, tantangan, dan sumber belajar. Kesalahpahaman adalah pemahaman konsep yang bertentangan dengan konsep ilmiah. Berdasarkan teori konstruktivis, kesalahpahaman adalah hasil dari struktur siswa yang berasal dari pengalaman, buku, guru, dan lingkungan (Suparno, 2013). Kesalahpahaman perlu segera dikenali dan diperbaiki. Jika dibiarkan, maka dapat mengganggu pemahaman siswa terhadap materi baru, mengakibatkan kinerja siswa yang buruk dan hasil belajar siswa yang rendah, sehingga tujuan pembelajaran tidak tercapai (Suwanto, 2013). Hal ini ditegaskan oleh (Noprianti & Utami, 2017) dimana miskonsepsi atau Kesalahpahaman dapat diidentifikasi dengan tes diagnostik.

Berdasarkan hasil observasi yang dilakukan oleh peneliti dilapangan yaitu di MAN 4 Jombang. menunjukkan bahwa pemahaman konsep siswa cukup baik. Hal ini ditunjukkan melalui hasil nilai pada mata pelajaran Fiqih dengan rata-rata 80. Factor yang mempengaruhi antara lain: a) metode pembelajaran yang digunakan guru hanya sekedar ceramah saja tanpa adanya variasi pembelajaran menggunakan media, b) siswa di dalam pembelajaran kurang antusias karena pembelajaran kurang menarik, c) evaluasi yang dilakukan oleh guru tidak berjalan dengan baik, d) kurangnya alat evaluasi yang dimiliki oleh guru untuk melakukan evaluasi terhadap pemahaman konsep pada siswa. Oleh karena itu, sistem pembelajaran dan evaluasi perlu diubah. Menurut (Fanani, 2018) dimana penilaian yang dikembangkan guru diharapkan dapat meningkatkan keterampilan berpikir yang lebih tinggi, meningkatkan kreativitas, dan membangun kemandirian siswa untuk pemecahan masalah.

Cara untuk mengetahui pemahaman konsep siswa adalah dengan menggunakan alat ukur yaitu tes diagnostik. Tes diagnostik adalah tes yang dirancang untuk mengidentifikasi kelemahan siswa, dan hasilnya dapat digunakan sebagai dasar untuk tindak lanjut sesuai dengan kelemahan siswa yang tepat. Selain untuk mengidentifikasi kelemahan, tes diagnostik pendidik adalah informasi yang dapat digunakan untuk memperbaiki proses pembelajaran, tetapi bagi siswa dapat digunakan untuk meningkatkan proses pembelajaran (Ismail et al., 2015).

Salah satu tes diagnostik yang bisa digunakan untuk mengidentifikasi kesulitan siswa dalam memahami konsep adalah tes *diagnostic test four tier berbasis high order thinking skills (hots)* atau diagnostik empat tingkat berdasarkan kemampuan penalaran tingkat tinggi, *Four-tier diagnostic test* biasanya digunakan untuk mengetahui dan mengidentifikasi level tentang memahami konsep pada siswa pada tes 4 tingkat. Yang membedakan antara tes pilihan 3 tingkat dengan 4 tingkat adalah adanya penambahan tingkat keyakinan terhadap alasan terhadap jawaban soal tersebut (Astuti et al., 2021).

Fiqih merupakan salah satu mata pelajaran Islam yang memperlakukan pengetahuan dan pemahaman hukum Islam baik yang berkaitan dengan ibadah maupun muamara sebagai pedoman hidup dan ibadah sehari-hari. Mata pelajaran fiqih termasuk hukum syariat agar ibadah yang dilakukan siswa atau peserta didik sesuai dengan hukum-hukum iyariat yang ada, sehingga ibadah yang dilakukan siswa jelas dapat digunakan untuk kebutuhan beribadah sehari-hari.

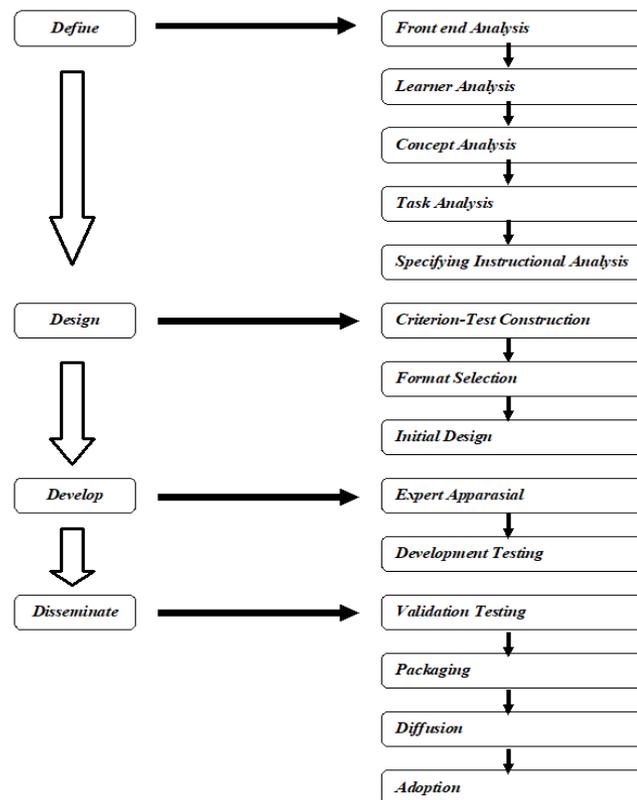
Peserta didik menemukan hal-hal yang berkaitan dengan amal ibadah dalam kehidupan sehari-hari sehingga perlu adanya siswa untuk memecahkan persoalan soal-soal dalam kategori berpikir tingkat tinggi (HOTS). Dengan adanya pengembangan instrument tes mata pelajaran fiqih ini sebagai upaya guru untuk mengenalkan problematika permasalahan pada mata pelajaran fiqih dengan kriteria soal berpikir tinggi (HOTS) yang perlu dipecahkan (Suryana, 2017). HOTS Keterampilan berpikir tingkat tinggi (*higher order thinking skills*) mencakup berpikir kritis, logis, reflektif, metakognitif, dan kreatif. Hal ini ditambahkan oleh (Sani, 2019) dimana siswa membutuhkan keterampilan berpikir kritis untuk memecahkan masalah dan mengambil keputusan. Mata pelajaran fiqih memerlukan pembuktian dengan menggunakan strategi untuk mengembangkan argumentasi sehingga penalaran erat kaitannya dengan kemampuan berpikir tinggi (HOTS) (Sa'idah & Isnaini, 2020).

Keterampilan berpikir merupakan gabungan dua kata yang memiliki arti berbeda, yaitu berpikir (*thinking*) dan keterampilan (*skills*). Berpikir adalah proses kognitif, yaitu mengetahui, mengingat, dan memahami, tetapi pentingnya keterampilan adalah pengumpulan dan pemilihan informasi, analisis, kesimpulan, penarikan ide, pemecahan masalah, mengevaluasi pilihan, membuat keputusan dan merefleksikan (Wilson, 2000).

Evaluasi hasil belajar diharapkan dapat membantu siswa meningkatkan kemampuan berpikir tinggi (HOTS) karena dapat mendorong siswa untuk berpikir secara luas dan mendalam tentang mata pelajaran. *Higher Order Thinking Skills* (HOTS) atau *Higher Order Thinking Skills* adalah bentuk revisi dari kata kerja operasional Bloom yang terdiri dari analisis (C4), evaluasi (C5), dan penciptaan (C6) yang digunakan untuk penyusunan soal. Guru perlu memiliki pengetahuan dan pengalaman untuk menunjang pekerjaannya sehingga dapat mengembangkan kemampuan berpikir tinggi siswanya (Aydin & Yilmaz, 2010).

METODE

Penelitian ini peneliti menggunakan metode penelitian pengembangan atau *Research and Development*. Dengan pendekatan pengembangan model 4D (*four-D model*) menurut (Thiagarajan et al., 1974). Yang terdiri dari 4 tahap yaitu, tahap pendefinisian (*define*), tahap perencanaan (*design*), tahap pengembangan (*develop*), dan tahap penyebaran (*disseminate*) (Gambar 1). Adanya penelitian ini untuk mengembangkan produk yang sudah ada, dikarenakan kurangnya alat atau instrument untuk mengevaluasi hasil belajar siswa. Penelitian ini diharapkan dapat menghasilkan produk instrument penilaian berupa *Diagnostic Test Four Tier* berbasis *High Order Thinking skills* terhadap penguasaan konsep siswa pada materi fiqih.



Gambar 1 Bagan Alur Model Pengembangan 4D

Dalam pelaksanaan kegiatan penelitian ini hanya berfokus pada 3 tahap yaitu, tahap pendefinisian (*define*), tahap perencanaan (*design*), dan tahap pengembangan (*develop*) (Thiagarajan et al., 1974). Pada penelitian ini hanya sampai pada pengembangan instrumennya saja, belum sampai pada penyebaran instrumen tersebut.

- Tahap pendefinisian (*Define*). Mengumpulkan data awal terkait kondisi awal di sekolah, kurikulum yang digunakan di sekolah, mengkaji siswa melalui observasi kepada siswa kelas 10 MIA MAN 4 Jombang, melakukan kajian Pustaka dengan menganalisis referensi penelitian terdahulu, dan menganalisis materi yang digunakan yaitu materi fiqih, serta mengintegrasikan hasil observasi kedalam rancangan produk pengembangan.
- Tahap Perencanaan (*Design*). Tahap ini untuk menyiapkan sebuah rancangan perangkat pembelajaran berupa instrument test yang akan dikembangkan. Didalam tahapan perencanaan bertujuan membuat rancangan instrument penilaian berupa *Diagnostic Test Four Tier* berbasis *High Order Thinking skills* terhadap penguasaan konsep siswa pada materi fiqih. Instrument penilaian ini dirancang dalam bentuk cetak sebanyak 16 soal pilihan ganda bertingkat 4 dimana tingkat 1 terkait soal, tingkat 2 berisi respon keyakinan terhadap jawaban ditingkat 1, tingkat 3 berisi alasan jawaban, dan tingkat 4 berisi respon keyakinan terhadap jawaban ditingkat 3.
- Tahap Pengembangan (*Develop*). Instrument diagnostic test ini melakukan penilaian yaitu 2 dosen dan 1 guru. Selain itu juga dilakukan uji respon kepada siswa terhadap instrument penilaiannya untuk mengetahui kevalidan dari tiap butir soal.

Jenis Data Kualitatif digunakan untuk menganalisis data yang diperoleh dari catatan, komentar, kritik, dan saran dari para ahli, dan praktisi kemudian digunakan untuk bahan revisi produk instrument. Sedangkan data kuantitatif berupa data nilai rata-rata dari kuesioner. Pengumpulan Data dengan Observasi dan penyebaran kuesioner. Observasi yang digunakan dalam penelitian ini adalah observasi lapangan, yang dipersiapkan secara sistematis tentang instrumen penilaian *diagnostic test four tier*. Sedangkan kuesioner yang digunakan pada penelitian ini adalah kuesioner untuk validasi produk pengembangan dan kuesioner untuk respon siswa.

Analisis data yang dihasilkan dari uji validasi oleh ahli, yaitu dosen, praktisi, atau guru mata pelajaran fiqih, selanjutnya dianalisis sebagai dasar untuk melakukan revisi produk. Analisis data dilakukan selama proses pengumpulan data dan setelah data diperoleh dalam satu tahapan yang telah dilakukan. Analisis sangat penting untuk menemukan data yang masih perlu dicari, pertanyaan yang

belum terjawab, dan kesalahan yang perlu diperbaiki berdasarkan pada rumus perhitungan dan kriteria kelayakan (Tabel 1) (Sugiyono, 2013).

$$X = \Sigma x / n$$

Keterangan:

- X = nilai rata-rata
- Σx = jumlah skor jawaban penilaian
- n = jumlah validator

Tabel 1 Kriteria validitas analisis rata-rata

Rata-rata	Kriteria Valid/Layak
3,28 – 4,00	Sangat Valid/Layak
2,52 – 3,27	Valid/Layak
1,76 – 2,51	Kurang Valid/Layak
1,00 – 1,75	Tidak Valid/Layak

Uji validitas dan uji reliabilitas instrumen penilaian ini juga dilakukan oleh siswa 20 peserta didik kelas X MA Negeri 3 Jombang. Suatu tes dapat dikatakan memiliki validitas yang tinggi apabila tes tersebut menjalankan fungsi ukurnya. Atau memberikan hasil ukur yang tepat dan akurat sesuai dengan maksud dikenakannya tes tersebut. Jika r hitung $\geq r$ tabel (uji 2 sisi dengan sig. 0,05), maka instrumen tersebut berkorelasi signifikan terhadap skor total (dinyatakan valid) sesuai Tabel 2.

Rumus korelasi *product moment* yang digunakan, yaitu: korelasi *product moment* dengan simpangan. Rumus korelasi *product moment* dengan simpangan:

$$r_{xy} = \frac{\Sigma xy}{\sqrt{(\Sigma x^2)(\Sigma y^2)}}$$

Keterangan:

- r_{xy} : Koefisien korelasi antara variabel x dan variabel y
- Σxy : Jumlah perkalian x dengan y
- x^2 : Kuadrat dari x
- y^2 : Kuadrat dari y

Tabel 2 Kriteria Penilaian Validitas Instrumen

Koefisien Validitas	Kategori
0,800 – 1,00	Sangat Tinggi
0,600 – 0,800	Tinggi
0,400 – 0,600	Cukup
0,200 – 0,400	Rendah
0,00 – 0,200	Sangat Rendah

Uji reliabilitas dilakukan dengan menggunakan rumus Kurder dan Richardson (KR 20). Rumus KR 20 ini merupakan rumus yang tepat untuk digunakan pada soal pilihan ganda dengan tingkat kesukaran yang berbeda. Berikut adalah rumus KR 20:

$$r_{11} = \left(\frac{k}{k-1} \right) \left(\frac{V_1 - \Sigma pq}{V_t} \right)$$

Keterangan:

- r_{11} : Reliabilitas instrumen
- k : Banyaknya butir soal
- V_t : Varians total
- p : Proporsi subjek yang menjawab betul pada suatu butir
- q : Proporsi subjek yang menjawab salah pada suatu butir

Adapun pedoman untuk memberikan interpretasi reliabilitas disajikan pada Tabel 3.

Tabel 3 Kriteria Reliabilitas Instrumen Penilaian

Koefisien Reliabilitas	Kategori
0,800 – 1,000	Sangat Tinggi
0,600 – 0,800	Tinggi
0,400 – 0,600	Sedang
0,200 – 0,400	Rendah
0,000 – 0,200	Sangat Rendah

HASIL DAN PEMBAHASAN

Pengembangan instrumen penilaian keterampilan berpikir kritis dalam hal ini mengacu pada model pengembangan Sugiyono yang terdiri dari 1) pendefinisian dengan melakukan analisis awal, analisis siswa, analisis tugas, analisis tugas dan spesifikasi tujuan; 2) desain produk diawali dengan menyusun kisi-kisi soal, menyusun soal, menyusun kunci jawaban dan validasi ahli; 3) pengembangan dimulai dari penilaian validasi ahli, uji coba terbatas dan uji coba luas.

- **Pendefinisian (*Define*)**

Tes diagnostik adalah tes yang digunakan siswa mengetahui kelemahan-kelemahan siswa sehingga hasil tersebut dapat digunakan sebagai dasar untuk memberikan tindak lanjut berupa perlakuan yang tepat sesuai dengan kelemahan yang dimiliki siswa (Depdiknas, 2007). Dalam penelitian ini, tes diagnostik dikembangkan dalam bentuk *four tier* dan digunakan untuk mengidentifikasi pemahaman konsep siswa pada materi fiqh.

Four tier diagnostic test (tes diagnostik empat tingkat) merupakan pengembangan dari *diagnostic test* pilihan ganda (*multiple-choice*) tiga tingkat (Arikunto, 2010).

- Tingkat pertama (*first tier*) atau disebut juga tingkatan jawaban (*answer tier*) merupakan tingkatan yang menyajikan pilihan jawaban dari pertanyaan/permasalahan yang diajukan.
- Tingkat kedua (*second tier*) merupakan tingkatan yang menyajikan berbagai skala keyakinan untuk mengukur seberapa yakin peserta didik menentukan dan memilih jawaban pada tingkat pertama.
- Tingkat ketiga (*third tier*) atau disebut juga tingkatan alasan (*reason tier*) merupakan tingkatan yang menyajikan alasan-alasan yang harus dipilih peserta didik terkait penentuan jawaban pada tingkat pertama.
- Tingkat keempat (*fourth tier*) tingkatan yang menyajikan berbagai skala keyakinan untuk mengukur seberapa yakin siswa menentukan dan memilih jawaban pada tingkat tiga.

- **Perancangan (*Design*)**

Tahap perancangan ini meliputi penyusunan instrumen, pemilihan format dan rancangan awal. Pada tahap ini merupakan tahap awal dalam pembuatan instrumen soal, yang dihasilkan kisi-kisi tes, instrumen penilaian yang dibuat dalam bentuk pilihan ganda (*multiple-choice*) dengan format *four tier* yang dihasilkan seperti pada Gambar 2.

Soal Instrumen Diagnostic Test Four Tier

1. Ozi merupakan anak dari orang yang berkecukupan dan anak yang baik, suatu saat Ozi mengetahui teman kelasnya ada yang tidak mempunyai tas sekolah yang layak untuk dipakai, sedangkan Ozi memiliki tas lebih dari satu. Karenanya Ozi memberikan salah satu tasnya kepada teman yang membutuhkan tersebut tanpa meminta imbalan apapun, apa yang dilakukan Ozi dinamakan

a. Sedekah
 b. Hadiah
 c. Hibah
 d. Jual
 e. Beli

Apakah anda yakin dengan jawaban anda?

a. Yakin
 b. Tidak

Alasan jawaban anda?

a. Memberikan sesuatu kepada orang lain dengan kemauan sendiri dengan tanpa adanya unsur paksaan melainkan secara ikhlas
 b. Memberikan sesuatu kepada orang lain dengan adanya imbalan
 c. Memberikan sesuatu kepada orang lain dengan terpaksa
 d. Memberikan sesuatu kepada orang lain yang menjadi tanggung jawabnya
 e. Saling bertukar barang

Apakah anda yakin dengan jawaban anda?

a. Yakin
 b. Tidak

Gambar 2. Contoh Soal dengan format *four tier diagnostic test*

• **Pengembangan (Develop)**

Dilakukan uji validitas oleh para ahli. Validasi ahli dilakukan oleh dua dosen PAI dan satu guru PAI. Validasi ahli dilakukan untuk menilai aspek petunjuk, aspek cakupan tes diagnostik dan aspek bahasa. Setelah melakukan validitas kepada para ahli dilakukan perbaikan pada kisi-kisi soal. Dari 25 soal yang disusun, terdapat 8 soal yang direvisi. Setelah dilakukan validasi ahli, dilakukan uji coba instrumen pada 20 siswa MA kelas XI di MA Negeri 3 Jombang. Hasil validasi kelayakan instrumen penilaian *diagnostic test four tier* pada kategori Layak dengan nilai 3,04. Analisis dari hasil uji coba tersebut terkait validitas dan reliabilitas soal. Tabel 4 menunjukkan hasil penilaian validasi ahli dari dua dosen PAI dan satu guru PAI.

Tabel 4 Hasil Penilaian Validasi Ahli

No.	Aspek Penilaian	Hasil Validator			Jumlah	Rata-rata
		V.1	V.2	V.3		
1.	Petunjuk	3	3	3	9	3,00
2.	Cakupan Tes Diagnostik	11	10	11	32	10,67
3.	Bahasa	13	15	13	41	13,67
Rata-rata Skor Total						9,11
Nilai kelayakan						3,04
Keterangan						Layak

Setelah dilakukan proses validasi kelayakan, selanjutnya dilakukan uji validitas instrumen dan uji reliabilitas instrumen. Uji validitas dan uji reliabilitas dilakukan oleh responden 20 peserta didik dengan mengisi jawaban pada instrumen penilaian *diagnostic test four tier*. Hasil uji validitas dihitung melalui

Fatati Nuriyana Mahmudah, Ino Angga Putra, Muhammad Aliyul Wafa, Andi Franoto
Rekonstruksi Diagnostic Test Four Tier Berbasis HOTS (High Order Thinking Skill)

SPSS 21 (*Analyze – Correlate – Bivariate*) menunjukkan bahwa ada 17 soal yang valid yaitu nomor soal 1, 2, 4, 5, 7, 10, 11, 14, 15, 16, 18, 19, 20, 21, 22, 23, 25 sebagai berikut.

		Correlations																												
		VAR00001	VAR00002	VAR00003	VAR00004	VAR00005	VAR00006	VAR00007	VAR00008	VAR00009	VAR00010	VAR00011	VAR00012	VAR00013	VAR00014	VAR00015	VAR00016	VAR00017	VAR00018	VAR00019	VAR00020	VAR00021	VAR00022	VAR00023	VAR00024	VAR00025				
VAR00001	Pearson Correlation Sig. (2-tailed)	1	.254	-.099	.254	.811	-.099	1.000	-.099	.034	.254	1.000	.034	-.099	.254	.811	.596	.050	.254	.254	.811	.596	.254	.254	.811	.596	1.000	.800		
VAR00002	Pearson Correlation Sig. (2-tailed)	.254	1	-.096	1.000	.313	-.096	.254	-.096	-.076	1.000	.254	.688	-.096	1.000	.313	.254	.459	1.000	1.000	.313	.254	1.000	1.000	1.000	1.000	-.115	.254	.810	
VAR00003	Pearson Correlation Sig. (2-tailed)	-.099	-.096	1	-.096	-.015	.608	-.099	.608	.327	-.096	-.099	-.140	1.000	-.096	-.015	-.099	.140	-.096	-.096	-.096	-.096	-.096	-.096	-.096	-.096	-.096	-.210	-.599	1.35
VAR00004	Pearson Correlation Sig. (2-tailed)	.254	1.000	-.096	1	.313	-.096	.254	-.096	-.076	1.000	.254	.688	-.096	1.000	.313	.254	.459	1.000	1.000	.313	.254	1.000	1.000	1.000	1.000	-.115	.254	.810	
VAR00005	Pearson Correlation Sig. (2-tailed)	.811	.313	-.015	.313	1	-.015	.811	-.015	.105	.313	1.000	.811	-.015	.313	1.000	.811	.157	.313	.313	.313	.313	.313	.313	.313	.313	.313	-.105	.811	.889
VAR00006	Pearson Correlation Sig. (2-tailed)	-.099	-.096	1	-.096	-.015	.608	-.099	.608	.327	-.096	-.099	-.140	1.000	-.096	-.015	-.099	.140	-.096	-.096	-.096	-.096	-.096	-.096	-.096	-.096	-.096	-.210	-.599	.081
VAR00007	Pearson Correlation Sig. (2-tailed)	1.000	.254	-.099	.254	.811	-.099	1	-.099	.034	.254	1.000	.034	-.099	.254	.811	.596	.050	.254	.254	.811	.596	.254	.254	.811	.596	.254	.254	.811	.800
VAR00008	Pearson Correlation Sig. (2-tailed)	-.099	-.096	1	-.096	-.015	.608	-.099	.608	.327	-.096	-.099	-.140	1.000	-.096	-.015	-.099	.140	-.096	-.096	-.096	-.096	-.096	-.096	-.096	-.096	-.096	-.210	-.599	.160
VAR00009	Pearson Correlation Sig. (2-tailed)	.034	-.076	.327	-.076	.185	-.140	.034	-.140	1	-.076	.034	-.111	.327	-.076	.185	.034	-.167	-.076	-.076	-.076	.105	.034	-.076	-.076	-.076	-.076	-.167	.034	.075
VAR00010	Pearson Correlation Sig. (2-tailed)	.254	1.000	-.096	1.000	.313	-.096	.254	-.096	-.076	1.000	.254	.688	-.096	1.000	.313	.254	.459	1.000	1.000	.313	.254	1.000	1.000	1.000	1.000	-.115	.254	.810	
VAR00011	Pearson Correlation Sig. (2-tailed)	1.000	.254	-.099	.254	.811	-.099	1.000	-.099	.034	.254	1	.034	-.099	.254	.811	.596	.050	.254	.254	.811	.596	.254	.254	.811	.596	.254	.254	.811	.800
VAR00012	Pearson Correlation Sig. (2-tailed)	.888	.688	-.140	.688	-.140	.688	-.140	.688	.327	-.111	.688	.034	1	-.140	.688	.185	.389	.288	.688	.688	.185	.389	.688	.688	.688	.688	.250	.688	.432
VAR00013	Pearson Correlation Sig. (2-tailed)	-.099	-.096	1.000	-.096	-.015	.608	-.099	.608	.327	-.096	-.099	-.140	1.000	-.096	-.015	-.099	.140	-.096	-.096	-.096	-.096	-.096	-.096	-.096	-.096	-.096	-.210	-.599	1.35
VAR00014	Pearson Correlation Sig. (2-tailed)	.281	1.000	-.096	1.000	.313	-.096	.254	-.096	-.076	1.000	.254	.688	-.096	1.000	.313	.254	.459	1.000	1.000	.313	.254	1.000	1.000	1.000	1.000	-.115	.254	.810	
VAR00015	Pearson Correlation Sig. (2-tailed)	.811	.313	-.015	.313	1.000	-.015	.811	-.015	.105	.313	1.000	.811	-.015	.313	1.000	.811	.157	.313	.313	.313	.313	.313	.313	.313	.313	.313	-.105	.811	.889
VAR00016	Pearson Correlation Sig. (2-tailed)	.596	.254	-.099	.254	.811	-.099	1.000	-.099	.034	.254	.596	.034	-.099	.254	.811	.596	.050	.254	.254	.811	.596	.254	.254	.811	.596	.254	.254	.811	.800
VAR00017	Pearson Correlation Sig. (2-tailed)	.050	.459	.140	.459	.157	.140	.050	.140	-.167	.459	.050	.250	.140	.459	.157	.050	1	.459	.459	.157	.050	.459	.459	.459	.459	.459	-.250	.050	.391
VAR00018	Pearson Correlation Sig. (2-tailed)	.254	1.000	-.096	1.000	.313	-.096	.254	-.096	-.076	1.000	.254	.688	-.096	1.000	.313	.254	.459	1.000	1.000	.313	.254	1.000	1.000	1.000	1.000	-.115	.254	.810	
VAR00019	Pearson Correlation Sig. (2-tailed)	.254	1.000	-.096	1.000	.313	-.096	.254	-.096	-.076	1.000	.254	.688	-.096	1.000	.313	.254	.459	1.000	1.000	.313	.254	1.000	1.000	1.000	1.000	-.115	.254	.810	
VAR00020	Pearson Correlation Sig. (2-tailed)	.811	.313	-.015	.313	1.000	-.015	.811	-.015	.105	.313	1.000	.811	-.015	.313	1.000	.811	.157	.313	.313	.313	.313	.313	.313	.313	.313	.313	-.105	.811	.889
VAR00021	Pearson Correlation Sig. (2-tailed)	.596	.254	-.099	.254	.811	-.099	1.000	-.099	.034	.254	.596	.034	-.099	.254	.811	.596	.050	.254	.254	.811	.596	.254	.254	.811	.596	.254	.254	.811	.800
VAR00022	Pearson Correlation Sig. (2-tailed)	.281	1.000	-.096	1.000	.313	-.096	.254	-.096	-.076	1.000	.254	.688	-.096	1.000	.313	.254	.459	1.000	1.000	.313	.254	1.000	1.000	1.000	1.000	-.115	.254	.810	
VAR00023	Pearson Correlation Sig. (2-tailed)	.254	1.000	-.096	1.000	.313	-.096	.254	-.096	-.076	1.000	.254	.688	-.096	1.000	.313	.254	.459	1.000	1.000	.313	.254	1.000	1.000	1.000	1.000	-.115	.254	.810	
VAR00024	Pearson Correlation Sig. (2-tailed)	.800	.610	.135	.610	.869	.081	.800	.180	.075	.610	.800	.432	.135	.610	.869	.761	.331	.610	.610	.869	.761	.331	.610	.610	.610	.610	-.010	.800	1
VAR00025	Pearson Correlation Sig. (2-tailed)	1.000	.254	-.099	.254	.811	-.099	1.000	-.099	.034	.254	1.000	.034	-.099	.254	.811	.596	.050	.254	.254	.811	.596	.254	.254	.811	.596	.254	.254	.811	.800

*. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Gambar 3 Hasil Uji Validitas melalui SPSS 21

Selain uji coba validitas, dilakukan juga uji reliabilitas instrumen. Instrumen dikatakan reliabel ketika instrumen tersebut dapat dengan ajeg memberikan data yang sesuai dengan kenyataan (Zulfikar et al., 2017). Hasil yang diperoleh dari perhitungan reliabilitas menggunakan SPSS 21 (*Analyze – Scale – Reliability Analysis*) adalah 0,901 dari 25 soal yang diuji cobakan kepada 20 peserta didik kelas X MA Negeri 3 Jombang. Hal ini menunjukkan bahwa nilai reliabilitas instrumen penilaian *diagnostic test four tier* ini pada kategori sangat tinggi. Hasil perhitungan reliabilitas menggunakan SPSS 21 sebagai berikut:

Case Processing Summary

		N	%
Cases	Valid	20	100.0
	Excluded ^a	0	.0
	Total	20	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.901	25

Gambar 4 Hasil Uji Reliabilitas melalui SPSS 21

SIMPULAN

Sesuai dengan hasil dan pembahasan yang telah dijabarkan diatas, maka diperoleh kesimpulan bahwa proses pengembangan instrumen penilaian hasil belajar terbagi menjadi dua tahap, yaitu tahap perancangan dan tahap uji coba. Pada tahap perancangan dihasilkan kisi-kisi tes, instrumen penilaian yang berupa soal *multiple-choice* yang terdiri dari 25 butir soal yang disusun berdasarkan *diagnostic test four tier*. Kelayakan instrumen penilaian yang dikembangkan dilihat dari tiga aspek yaitu, aspek petunjuk, aspek cakupan tes diagnostik dan aspek bahasa dengan nilai rata-rata 3,04 pada kategori layak. Setelah dilakukan uji kelayakan maka dilanjutkan uji validitas dan uji reliabilitas kepada responden sejumlah 20 siswa yang dianalisis melalui SPSS 21 dengan hasil: a) ada 17 butir soal yang valid dan b) nilai reliabilitas seluruh butir soal sangat tinggi dengan nilai 0,901.

Terima kasih peneliti sampaikan kepada pihak-pihak yang telah membantu dalam pelaksanaan kegiatan penelitian kami yaitu: 1) Lembaga Penelitian dan Pengabdian pada Masyarakat (LPPM) Universitas KH. A. Wahab Hasbullah (UNWAHA), 2) Sivitas Akademika Program Studi Pendidikan Agama Islam, dan 3) seluruh keluarga peneliti yang juga mendoakan kelancaran kegiatan penelitian ini.

DAFTAR RUJUKAN

- Agustin, A., Supardi, K. I., & Sunarto, W. (2018). Pengaruh Penggunaan Peta Konsep Berbasis Multilevel terhadap Pemahaman Konsep Kimia SMA. *Chemistry in Education*, 7(2), 8–13.
- Arikunto, S. (2010). Prosedur Penelitian Pendekatan Praktik Edisi Revisi VI. In *Rineka Cipta*.
- Astuti, I. A. D., Bhakti, Y. B., & Prasetya, R. (2021). Four Tier-Magnetic Diagnostic Test (4T-MDT): Instrumen Evaluasi Medan Magnet Untuk Mengidentifikasi Miskonsepsi Siswa. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, 5(2), 110–115. <https://doi.org/10.30599/jipfri.v5i2.1205>
- Aydin, N., & Yilmaz, A. (2010). The Effect of Constructivist Approach in Chemistry Education on Students ' Higher Order Cognitive Skills. *Journal of Education*, 39(2), 57–68.
- Depdiknas. (2007). *Tes Diagnostik*. Direktorat Pembinaan Sekolah Menengah Pertama.
- Fanani, M. Z. (2018). Strategi Pengembangan Soal Hots Pada Kurikulum 2013. *Edudeena*, 2(1), 57–76. <https://doi.org/10.30762/ed.v2i1.582>
- Ismail, I. , Samsudin, A., Suhendi, E., & Kaniawati, I. (2015). Diagnostik Miskonsepsi Melalui Listrik Dinamis Four Tier Test. *Simposium Nasional Inovasi Dan Pembelajaran Sains*.
- Noprianti, E., & Utami, L. (2017). Penggunaan Two-Tier Multiple Choice Diagnostic Test Disertai Cri Untuk Menganalisis Miskonsepsi Siswa. *JTK (Jurnal Tadris Kimiya)*, 2(2), 124–129. <https://doi.org/10.15575/jtk.v2i2.1876>
- Putra, I. A., Sujarwanto, E., & Pertiwi, N. A. S. (2018). Analisis pemahaman konseptual mahasiswa pada materi kinematika partikel melalui tes diagnostik. *Jurnal Riset Dan Kajian Pendidikan Fisika*, 5(1), 10. <https://doi.org/10.12928/jrpkpf.v5i1.8923>
- Rohmah, Z.-, Handika, J.-, & Huriawati, F.-. (2019). E-Diagnostic Test untuk Mengungkap Miskonsepsi Kinematika. *SPEKTRA: Jurnal Kajian Pendidikan Sains*, 5(1), 22. <https://doi.org/10.32699/spektra.v5i1.86>
- Sa'idah, N., & Isnaini, I. K. (2020). PENGEMBANGAN INSTRUMEN TES HIGHER ORDER THINKING SKILL (HOTS) MATA PELAJARAN FIKIH KELAS XI DI MA MASALIKIL HUDA TAHUNAN JEPARA. *Jurnal Pendidikan Islam*, 11(2), 209–220. <https://media.neliti.com/media/publications/367415-pengembangan-instrumen-tes-higher-order-f6b2bb92.pdf>
- Sani, R. A. (2019). *Pembelajaran Berbasis HOTS (Higher Order Thinking Skill)*. Tira Smart.
- Setiawan. (2008). *Konsep Dan Makna Pembelajaran*. Alfabeta.
- Siswaningsih, W. (2014). Pengembangan Tes Diagnostik Two-Tier Untuk Mengidentifikasi Miskonsepsi Pada Materi Kimia Siswa Sma. *Jurnal Pengajaran Matematika Dan Ilmu Pengetahuan Alam*, 19(1), 117. <https://doi.org/10.18269/jpmipa.v19i1.487>
- Sugiyono. (2013). Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D Sugiyono. 2013. "Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D." Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D. <https://doi.org/10.1>. In *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D*. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Suparno, P. (2013). *Miskonsepsi dan Perubahan Konsep dalam Pendidikan Fisika*. PT. Gramedia Widiasarana Indonesia.

- Suryana, E. (2017). Pengembangan Bahan Ajar Fiqih Dengan Menggunakan Model Pembelajaran Gagne dan Briggs Berbasis Flipbook di MTS N Panca Mukti Kelas VIII Kecamatan Pondok Kelapa Kabupaten Bengkulu Tengah. *An-Nizom, Vol.2(2)*, h. 305-312.
- Suwarto. (2013). *Pengembangan Tes Diagnostik dalam pembelajaran, panduan praktis bagi pendidik dan calon pendidik*. Pustaka Pelajar.
- Thiagarajan, S., Dorothy S., S., & Melvyn I., S. (1974). *Instructional development for training teachers of exceptional children*. Indiana University.
- Ulva, Y. ., Santosa, A., & Parlan. (2016). Identifikasi tingkat pemahaman konsep larutan penyangga. *Jurnal Pembelajaran Kimia (J-PEK)*, 1(2), 69–75.
- Wilson, V. (2000). Education forum on teaching thinking skills. In *Edinburgh Report, the scottish council for research in education*.
- Zulfikar, A., Samsudin, A., & Saepuzaman, D. (2017). Pengembangan Terbatas Tes Diagnostik Force Concept Inventory Berformat Four-Tier Test. *WaPFI (Wahana Pendidikan Fisika)*, 2(1), 43–49. <https://doi.org/10.17509/wapfi.v2i1.4903>