

ANALISIS KUALITAS SOAL THREE TIER TEST MENGGUNAKAN TEORI RESPON BUTIR KLASIK

Anas Irwan, Ahmat Purwadi Saleh Kerans

Pendidikan Fisika Fakultas Tarbiyah dan Keguruan UIN Alauddin Makassar, anas.irwan@uin-alauddin.ac.id.

Abstrak

Penelitian dilakukan dengan tujuan untuk mengetahui kualitas soal three tier test ujian akhir semester mata kuliah Fisika Dasar 1 menggunakan teori respon butir klasik pada mahasiswa Prodi Pendidikan Fisika Fakultas Tarbiyah dan Keguruan UIN Alauddin Makassar. Jenis penelitian ini adalah mix methode dengan desain penelitian multiphase design. Penelitian ini dilakukan di Prodi Pendidikan Fisika Fakultas Tarbiyah dan Keguruan UIN Alauddin Makassar. Hasil penelitian menunjukkan bahwa uji validitas isi yang dilakukan dengan uji Gregory menunjukkan tingkat validnya sebesar 0,9 dan uji validitas konstruk diperoleh 28 butir yang valid dan 12 butir yang tidak valid, sedangkan uji reliabilitas menggunakan Alpha Cronbach diperoleh 0,6045 (cukup reliabel). Dari hasil analisis kualitas soal diperoleh sebanyak 11 butir baik, sebanyak 20 butir cukup dan sebanyak 9 butir tidak baik. Impilkasi dari pelitian ini adalah soal yang mempunyai kualitas baik dan cukup baik dapat dijadikan sebagai bank soal, sedangkan butir soal yang tidak baik sebaiknya dilakukan revisi.

Kata Kunci: Kualitas Soal; Three Tier Test; Pemahaman Konsep; Miskonsepsi

Abstract

This study aims to determine the quality of the final three-semester level test of Basic Physics 1 course using classical item response theory for students of the Physics Education Study Program, Faculty of Tarbiyah and Teacher Training UIN Alauddin Makassar. This type of research is a mix method with a multiphase design research design. This research was conducted at the Physics Education Study Program, Faculty of Tarbiyah and Teacher Training at UIN Alauddin Makassar. The results showed that the content validity test performed with the Gregory test showed a validity level of 0.9 and the construct validity test obtained 28 valid items and 12 valid items, while the reliability test using Cronbach's Alpha obtained 0.6045 (reliable enough). From the results of the analysis of the quality of the questions obtained as many as 11 items are good, as many as 20 items are sufficient and as many as 9 items are not good. The implication of this research is that questions that have good quality and are good enough can be used as question banks, while items that are not good must be revised.

Keywords: Question Quality; Three Tier Test; Concept Understanding; Misconception

PENDAHULUAN

Dalam Undang-undang Sisdiknas Tahun 2003 Pasal 3, menyatakan bahwa fungsi dari pendidikan nasional adalah pengembangan dan pembentukan watak serta menjadikan bangsa yang bermartabat dan mempunyai peradaban dalam mencetuskan mencerdaskan kehidupan bangsa”.

Evaluasi merupakan komponen yang tidak kalah pentingnya dan tahap yang dilalui oleh seorang dalam mengetahui keefektifan proses pembelajaran yang telah dilakukan. Hasil yang telah didapatkan dari evaluasi akan dijadikan sebagai umpan balik (*feed-back*) bagi seorang

dalam perbaikan dan penyempurnaan proses pembelajaran (Arifin Zainal, 2013). Salah satu cara untuk melaksanakan evaluasi adalah dengan menggunakan tes. Tes adalah seperangkat rangsangan (stimulus) yang digunakan untuk mendapatkan jawaban dari responden untuk dijadikan sebagai bahan ditetapkan standar skor. Tes biasanya terdiri atas tes objektif dan esai (Uno & Koni, 2013). Tes yang keperluan untuk menjawab semua informasi mengenai tes adalah tes objektif (Purwanto, n.d.)

Three tier test merupakan cara mengevaluasi pengetahuan konsep serta

miskonsepsi kuantitas mahasiswa dalam bentuk tes diagnosis. *Three tier test* memiliki tiga tingkatan, yaitu tingkat pertama *content tier* adalah menanyakan pengetahuan mahasiswa tentang konsep dari pilihan ganda. Tingkatan kedua, *reason tier* adalah penalaran mahasiswa dari proses menjawab pada tingkatan pertama. Tingkatan ketiga, *certainty respon index* adalah pertanyaan mengenai keyakinan mahasiswa tentang jawaban tingkatan pertama (Arslan et al., 2012). Sehingga mahasiswa tersebut tidak langsung memutuskan untuk menjawab dengan cara menebak pada tingkat pertama dan kedua (Caleon & Subramaniam, 2010)

Berdasarkan proses wawancara yang telah peneliti lakukan terhadap dosen pengampu mata kuliah Fisika Dasar didapatkan bahwa selama ini soal yang digunakan pada ujian akhir semester adalah berbentuk essay. Alasan dosen pengampu mata kuliah memilih soal dalam bentuk essay karena dapat dilihat kemampuan mahasiswa melalui proses-proses dalam menyelesaikan soal tersebut. Setiap proses-proses dalam penyelesaian soal tersebut mempunyai poin-poin sampai hasil akhir yang didapatkan oleh mahasiswa.

Oleh karena itu, perlu dilakukan analisis kualitas terhadap soal ujian akhir semester mahasiswa yang dibuat oleh dosen tersebut. Hal ini penting dilakukan sebagai bahan evaluasi dosen terhadap mahasiswa dan sebagai jaminan untuk kualitas soal yang telah disusun. Hal tersebut sesuai juga dengan hasil wawancara kepada Ketua Program Studi Pendidikan Fisika yang mengatakan “bahwa belum pernah adanya pengontrolan terhadap kualitas soal yang disusun oleh dosen. Dari permasalahan tersebut diketahui bahwa perlu dilakukan analisis butir soal. Selain hal tersebut, terdapat juga soal yang diambil dari buku, dan soal ujian akhir semester dari tahun sebelumnya yang tidak dijamin kualitas soalnya. Oleh karena itu, pada penelitian yang difokuskan adalah kualitas soal *three tier test* yang digunakan yaitu pada analisis butir klasik bukan pada penguasaan konsep maupun miskonsepsi.

Berdasarkan hasil wawancara tersebut, maka peneliti tertarik untuk melakukan penelitian dengan judul “*Analisis Kualitas Soal Three Tier Test Ujian Akhir Semester Mata Kuliah Fisika Dasar 1 Menggunakan Teori Respon Butir Klasik Pada Mahasiswa Prodi Pendidikan Fisika Fakultas Tarbiyah dan Keguruan UIN Alauddin Makassar*”.

Sehingga dalam penelitian ini, kegiatan analisis kualitas soal ini sangat penting dalam upaya mendapatkan instrumen dengan penilaian yang berkategori lebih bermutu. Melalui kegiatan analisis kualitas soal dapat diketahui butir-butir soal yang masuk ke dalam kategori sangat baik, baik, kurang baik, cukup dan buruk. Dari kegiatan analisis kualitas soal ini dapat diperoleh sebagai bahan informasi mengenai bermutu atau tidaknya dari soal, sekaligus sebagai petunjuk dalam melakukan revisi soal yang telah disusun. Kegiatan menganalisis atau menelaah kualitas diharapkan dapat menghasilkan soal-soal yang lebih bagus untuk digunakan pada evaluasi pembelajaran (Peşman & Eryilmaz, 2010)

Menurut (Sudjana, 2009), bahwa analisis kualitas soal adalah suatu langkah yang disusun secara terstruktur, untuk menelaah butir tes yang telah disusun dan dijadikan bahan acuan atau standar dalam menyusun soal mendatang. Tujuan dari dilakukan penelaahan kualitas tes adalah untuk mengkaji dan mengetahui lebih dalam setiap soal agar diperoleh soal yang lebih bagus dan lebih bermutu untuk digunakan pada saat melakukan diagnosa pemahaman konsep. Disamping itu tujuan dari analisis kualitas soal adalah sebagai bahan deskripsi informasi yang tepat sesuai dengan tujuan dari soal tersebut dibuat.

Tujuan utama dari penelaahan kualitas soal dalam sebuah tes yang dibuat dosen adalah untuk mengetahui kekurangan-kekurangan apa saja yang ada dalam soal yang disusun atau dalam proses pembelajaran. Berdasarkan tujuan tersebut, aktifitas telaah kualitas soal memiliki banyak didapat manfaatnya, antara lain yaitu (1) menentukan apakah suatu fungsi butir soal sesuai

dengan yang diinginkan, (2) sebagai bahan masukan dan diskusi didalam melakukan pemebelajaran dikelas, (3) sebagai masukan tambahan kepada pengajar tentang kesulitan yang dihadapi mahasiswa dalam kelas, (4) memberi masukan pada bagian tertentu guna pengembangan kurikulum, (5) memberikan perbaikan pada materi yang diukur atau dinilai, (6) meningkatkan kemampuan dalam penyusunan soal (Djanuarsih, 2012)

Analisis item butir soal pada umumnya dilakukan melalui dua cara, yaitu analisis kualitatif dan analisis kuantitatif. Aspek yang diperhatikan dalam penelaahan secara kualitatif mencakup aspek materi, konstruksi, bahasa atau budaya, dan kunci jawaban. Analisis item butir soal secara kuantitatif merupakan penelaahan butir soal didasarkan pada bukti empirik (Popham, 2010).

Ada dua pendekatan dalam analisis secara kuantitatif, yaitu pendekatan secara klasik dan modern. Menurut Mamun Ali Naji Qasem, analisis butir soal klasik (*classical test theor* menyatakan bahwa: *classical test theory introduces three concepts test score, true score, and error score. Within that theoretical framework, models of various forms have been formulated.* Artinya: teori tes klasik memperkenalkan tiga konsep yaitu skor tes, skor benar, dan skor salah. Dalam kerangka teoritis itu, model berbagai bentuk telah dirumuskan (Qasem, 2013)

Sedangkan, *item response theory* (IRT) adalah sebuah model probabilitas yang berusaha menjelaskan hubungan antara respon seseorang terhadap sebuah butir dengan variabel laten (kemampuan/*ability* atau sifat/*trait*) yang diukur oleh tes tersebut. Dalam hal ini, respon atau kinerja peserta tes merupakan hal yang dapat diamati (*observable*) sedangkan sifat atau kemampuan merupakan sesuatu yang tidak tampak (*unobservable*) yang mendasari kinerja pada tes tersebut (Embretson & Reise, 2000)

Menurut (Maulini et al., 2016) bahwa *three tier test* adalah salah satu jenis tes diagnostik yang

menggunakan identifikasi miskonsepsi dan pemahaman konsep mahasiswa. *Three tier test* memiliki tiga tingkatan, pertama adalah menanyakan pengetahuan mahasiswa tentang konsep dari pilihan ganda. Tingkatan kedua adalah penalaran mahasiswa dari proses menjawab pada tingkatan pertama. Tingkatan ketiga adalah pertanyaan mengenai keyakinan mahasiswa tentang jawaban tingkatan pertama dan kedua.

Test ini dianggap agak lebih efisien dari test pilihan ganda, sebab terdiri atas dua tingkatan; pada tingkat pertama *content tier* yang mengukur pengetahuan responden terkait suatu konsep atau materi dan pada tingkat kedua *reason tier* untuk melihat alasan dibalik jawaban yang diberikan oleh responden pada *content tier*. Dengan adanya soal pada tingkat kedua dapat dilihat apakah jawabannya benar atau salah yang diberikan oleh responden, dimana jika jawabannya di tingkat pertama benar maka responden dianggap paham terhadap konsep dari materi yang telah diajarkan akan tetapi apabila jawabannya salah maka responden tersebut dianggap tidak paham terhadap konsep yang telah diajarkan atau disebut miskonsepsi. Meski begitu, jenis test ini dianggap masih memiliki kekurangan dimana tidak mampu membedakan kesalahan dari responden karena kurangnya pengetahuan dengan kesalahan karena sebuah miskonsepsi, dan juga, tidak bisa membedakan responden yang memberikan jawaban dengan menebak atau memang paham dengan konsep dari suatu materi (Popham, 2010)

Adapun beberapa faktor yang mempengaruhi pemahaman konsep tersebut, dikarenakan kurangnya perhatian saat pembelajaran berlangsung ataupun waktu belajar, pemahaman konsep yang sekarang, juga dipengaruhi oleh pemahaman sebelumnya, karena kita ketahui bersama, bahwa ilmu fisika memiliki saling keterkaitan dalam materinya, ataupun dipengaruhi oleh tenaga pendidik serta sarana yang digunakan sebelumnya (Irwan, 2015)

Dalam penelitian metode tes diagnostik jenis *three tier test* digunakan untuk mengukur miskonsepsi mahasiswa. Jenis tes ini kemungkinan jawabannya adalah seperti pada tabel berikut:

Tabel 1. Kombinasi jawaban soal *three tier test*

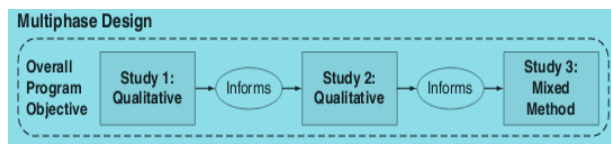
Tingkat soal	Tipe Jawaban	Kategori	Kode
<i>Three Tier Test</i>	B + B+ Yakin	Paham Konsep	PK
	B + B+ Tidak Yakin	Kurang Paham Konsep	KPK
	S + S + Tidak Yakin	Tidak Paham Konsep	TPK
	S + B + Tidak Yakin	Menebak	M
	B + S + Tidak Yakin		
	B + S + Yakin		
	S + S + Yakin	Miskonsepsi	MK
	S + B + Yakin		

(Sumber : Suhendi, dkk, 2014)

Oleh karena itu tujuan tujuan dari penelitian ini untuk mengetahui kualitas soal *three tier test* ujian akhir semester mata kuliah Fisika Dasar 1 menggunakan teori respon butir klasik pada mahasiswa Prodi Pendidikan Fisika Fakultas Tarbiyah dan Keguruan UIN Alauddin Makassar dengan manfaat sebagai bahan evaluasi bagi dosen pengampuh mata kuliah dalam menyusun soal kedepannya.

METODE PENELITIAN

Jenis penelitian ini adalah *mix metode*, yaitu penelitian dilakukan dengan prosedur untuk mengumpulkan, menganalisis dan mencampur metode kuantitatif dan kualitatif dalam suatu penelitian atau serangkaian penelitian untuk memahami penelitian (Creswell, 2012), dengan desain penelitian adalah *multiphase design*.



Gambar 1: Desain penelitian *multiphase design*

Subjek dalam penelitian ini adalah dosen mata kuliah fisika dasar dan 76 orang mahasiswa. Teknik pengambilan subjek pada penelitian ini dilakukan dengan teknik *non-probability sampling* yaitu dengan menggunakan teknik *convenience sampling*. Dalam *teknik convenience sampling* (kemudahan), peneliti hanya memilih peserta karena mau dan bersedia untuk dijadikan objek penelitian (Creswell, 2012)

HASIL DAN PEMBAHASAN

1. Hasil penelitian

Hasil penelitian yang diperoleh melalui penelitian yang dilakukan oleh peneliti pada tanggal 18 Desember 2019 di Program Studi (Prodi) Pendidikan Fisika Fakultas Tarbiyah dan Keguruan UIN Alauddin Makassar. Teknik pengumpulan data yang digunakan adalah teknik dokumentasi. Hasil penelitian setelah dilakukan analisis sebagai berikut.

a. Tingkat 1 (*content tier*)

1) Validitas

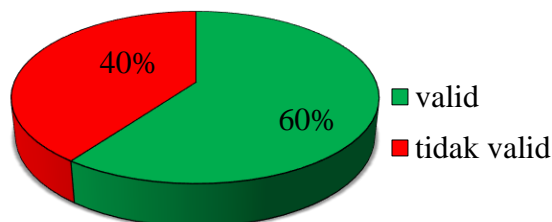
Validitas merupakan produk dari validasi. Validasi adalah suatu proses yang dilakukan oleh penyusun atau pengguna instrument untuk mengumpulkan data secara empiris guna mendukung kesimpulan yang dihasilkan oleh skor instrument. Sedangkan validitas adalah kemampuan suatu alat ukur untuk mengukur sasaran ukurnya (Amiriono & Daryanto, 2016)

a) Validitas isi

Berdasarkan hasil dari uji validasi yang dilakukan dengan uji Gregory menunjukkan tingkat validnya sebesar 0,9. Uji Gregory memperbolehkan untuk menguji validitas secara keseluruhan dari instrument tes sehingga dapat dikatakan bahwa soal tersebut sangat valid untuk digunakan.

b) Validitas konstruk

Berdasarkan hasil uji validitas konstruk menunjukkan bahwa dari 40 butir soal yang digunakan, terdapat 24 butir (60 %) soal valid dan 16 butir (40 %) soal yang tidak valid.



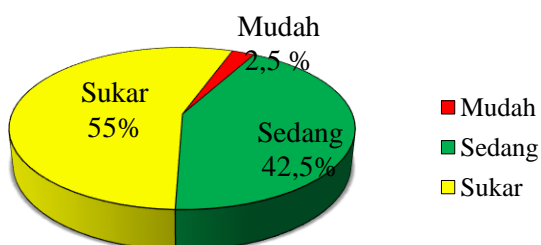
Gambar 2: diagram pie uji validitas soal tingkat 1 (*content tier*)

2) Reliabilitas

Berdasarkan hasil analisis reliabilitas menggunakan *Ms. Excel 2010* diperoleh nilai $r_{hitung} = 0,6045$. Berdasarkan r_{tabel} acuan penentuan reliabilitas data dengan syarat $r_{hitung} > r_{tabel}$, sehingga soal tersebut termasuk cukup reliabel.

3) Tingkat kesukaran

Berdasarkan uji tingkat kesukaran dengan menggunakan *Ms. Excel 2010* menunjukkan bahwa dari 40 butir soal pilihan ganda, dimana soal yang termasuk kategori mudah sebanyak 1 butir (2,5 %), soal yang termasuk dalam kategori sedang sebanyak 17 butir (42,5 %) dan sedangkan yang masuk kategori sukar sebanyak 22 butir (55 %).

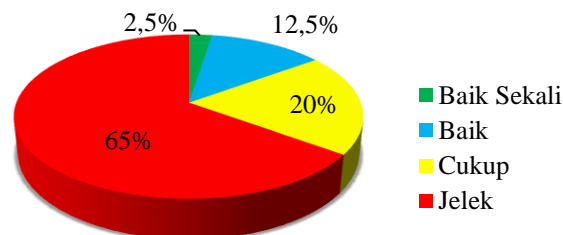


Gambar 3: diagram pie tingkat kesukaran soal tingkat 1 (*content tier*)

4) Daya pembeda

Berdasarkan hasil analisis daya pembeda menggunakan *Ms. Excel 2010* menunjukkan bahwa sebanyak 1 butir (2,5 %) soal memiliki daya pembeda baik sekali, 5 butir (12,5 %) soal memiliki daya pembeda baik, 8 butir (20 %) soal

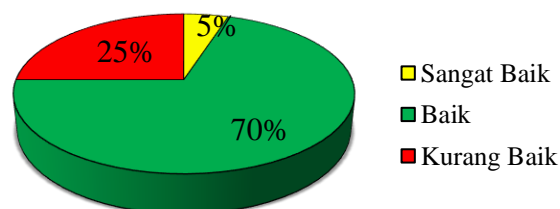
memiliki daya pembeda cukup dan 26 butir (65 %) soal memiliki daya pembeda jelek.



Gambar 4: diagram pie daya pembeda soal tingkat 1 (*content tier*)

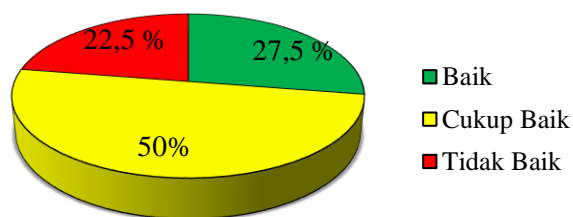
5) Efektivitas pengecoh

Berdasarkan hasil analisis daya pembeda menggunakan *Ms. Excel 2010* menunjukkan bahwa 2 butir (5 %) soal memiliki kualitas sangat baik, terdapat juga 28 butir (70 %) soal yang masuk dalam kategori baik, terdapat 10 butir (25 %) soal kurang baik, untuk butir soal yang masuk pada kategori buruk dan sangat buruk tidak ada.



Gambar 5: diagram pie efektivitas pengecoh soal tingkat 1 (*content tier*)

Dari hasil analisis yang diperoleh dengan menggunakan pendekatan teori tes klasik terdiri atas validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh soal *three tier test* (tes tiga tingkat) menunjukkan bahwa 11 butir soal memiliki kategori baik, terdapat juga 20 butir soal yang masuk dalam kategori baik, sedangkan soal yang masuk kategori tidak baik terdapat 9 butir soal.



Gambar 6: diagram pie kategori kualitas soal tingkat I (*content tier*)

b. Tingkat II (*Reason Tier*)

Berdasarkan hasil pemberian alasan yang lebih dominan pada soal tingkat II (*reason tier*) menunjukkan bahwa terdapat 15 butir (37,5 %) soal yang lebih dominan menjawab benar dan terdapat 25 butir (62,5 %) soal yang lebih dominan menjawab salah.

c. Tingkat III (*Certainty Respon Index*)

Berdasarkan hasil pemberian yakin dan tidak yakin yang lebih dominan pada soal tingkat III (*certainty respon index*) terdapat 20 butir (50 %) soal yang lebih dominan menjawab yakin dan terdapat 20 butir (50%) soal yang lebih dominan menjawab tidak yakin.

d. Analisis Kombinasi Jawaban Soal *Three Tier Test*

Three tier test merupakan instrumen dengan tiga tingkatan; *content tier* yang mengukur pengetahuan responden terkait suatu konsep/materi, *reason tier* untuk melihat alasan dibalik jawaban yang diberikan oleh responden pada *content tier*, dan *certainty respon index* yang mengukur seberapa percaya diri responden akan jawabannya di tingkat pertama dan kedua.

Berdasarkan hasil analisis pola kombinasi jawaban mahasiswa dari 40 butir soal *three tier test* menunjukkan bahwa terdapat 21 butir soal yang lebih dominan mahasiswa tidak paham konsep, terdapat 3 butir soal yang lebih dominan mahasiswa kurang paham konsep, terdapat 19 butir soal yang lebih dominan mahasiswa menebak, terdapat 21 butir soal yang lebih dominan mahasiswa miskonsepsi, dan terdapat 18 butir soal yang lebih dominan mahasiswa paham konsep.

2. Pembahasan

Dari hasil analisis yang diperoleh dengan menggunakan pendekatan teori tes klasik terdiri atas tingkat kesukaran, daya pembeda, dan efektivitas pengecoh soal *three tier test* (tes tiga tingkat) tingkat I (*content tier*) ujian akhir semester (UAS) mata kuliah Fisika Dasar 1 pada mahasiswa semester satu Program Studi (Prodi) Pendidikan Fisika Fakultas Tarbiyah dan Keguruan UIN Alauddin Makassar menunjukkan bahwa terdapat 11 butir soal memiliki kategori baik, terdapat juga 20 butir soal yang masuk dalam kategori baik, sedangkan soal yang masuk kategori tidak baik terdapat 9 butir soal.

Pada soal *three tier test* (tes tiga tingkat) tingkat II (*reason tier*) ujian akhir semester (UAS) mata kuliah Fisika Dasar 1 pada mahasiswa semester satu Program Studi (Prodi) Pendidikan Fisika Fakultas Tarbiyah dan Keguruan UIN Alauddin Makassar berdasarkan hasil penelitian yang diperoleh dapat diketahui bahwa masih banyak memberikan alasan yang salah terutama pada soal yang berada pada ranah kognitif C1, C2, C3, dan C4. Sedangkan tingkat III (*certainty respon index*) ujian akhir semester (UAS) mata kuliah Fisika Dasar 1 pada mahasiswa semester satu Program Studi (Prodi) Pendidikan Fisika Fakultas Tarbiyah dan Keguruan UIN Alauddin Makassar berdasarkan hasil penelitian yang diperoleh dapat diketahui bahwa masih banyak yang memberikan keyakinan yang kurang terutama pada soal yang berada pada ranah kognitif C3 dan C4.

Beberapa penelitian mengenai analisis kualitas soal dan *three tier test* yang mendukung adalah (Natalia et al., 2017) yang mengatakan bahwa pada saat melakukan tes kepada peserta didik pada awal permulaan tes dimulai peserta didik masih mengerjakan tes dengan tenang akan tetapi pada soal-soal terakhir tes terdapat beberapa mahasiswa yang melakukan contekan. Hal ini dapat dikatakan bahwa peserta didik tersebut pada soal tingkat yang ketiga mahasiswa tidak memahami konsep fisika secara baik.

Sedangkan menurut (Prawira, 2018) mengatakan bahwa adanya miskonsepsi pada mahasiswa calon guru fisika yaitu pada konsep kinematika partikel. Mahasiswa banyak yang mengalami miskonsepsi pada subkonsep tentang GLBB dan lintasan benda pada gerak peluru. Faktor penyebab terjadinya miskonsepsi berasal dari diri mahasiswa calon guru fisika sendiri yaitu pemikiran asosiatif, reasoning yang tidak lengkap/salah, dan intuisi yang salah.

Sedangkan menurut (Bunawan, 2014) menyatakan bahwa rata-rata skor pengetahuan konseptual responden yang diuji dengan menggunakan tes tiga tingkat tidak ada masalah terhadap pemahaman konsep responden dilihat dari beda signifikan terhadap skor rata-rata uji.

Dari beberapa pemaparan penelitian di atas dapat diketahui bahwa untuk mengetahui tingkat pemahaman konsep pada mahasiswa dapat dilakukan dengan menggunakan soal *three tier test* yang terlebih dahulu dilakukan analisis kualitas soal.

KESIMPULAN

Berdasarkan hasil penelitian diatas dapat disimpulkan bahwa: 1) Kualitas butir soal *three tier test* (tes tiga tingkat) tingkat I (*content tier*) setelah dilakukan analisis diketahui bahwa terdapat 11 butir soal memiliki kategori baik, terdapat 20 butir (50 %) soal yang masuk dalam kategori baik, dan soal yang masuk kategori tidak baik terdapat 9 butir soal. 2) Gambaran pemberian alasan soal *three tier test* (tes tiga tingkat) tingkat II (*reason tier*) yang diperoleh dapat diketahui bahwa masih banyak yang memberikan alasan yang salah terutama pada soal yang berada pada ranah kognitif C3 dan C4. Hal ini menandakan kurangnya penguasaan konsep pada materi yang telah diajarkan. 3) Pada soal *three tier test* (tes tiga tingkat) tingkat III (*certainty respon index*) yang diperoleh dapat diketahui bahwa masih banyak yang memberikan keyakinan yang kurang terutama pada soal yang berada pada ranah kognitif C3 dan C4.

DAFTAR PUSTAKA

- Amirono, & Daryanto. (2016). *Evaluasi dan penilaian pembelajaran Kurikulum 2013*. Gava Media.
- Arifin Zainal. (2013). *Evaluasi Pembelajaran*. Remaja Rosdakarya.
https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Arifin%2C+Zainal%2C++2013%2C+Evaluasi+Pembelajaran%2C+Remaja+Rosdakarya%2C+Bandung.&btnG=#d=gs_cit&t=1659797664052&u=%2Fscholar%3Fq%3Dinfo%3AkWL7UC62ITEJ%3Ascholar.google.com%2F%26output%3Dcite%26sci
- Arslan, H. O., Cigdemoglu, C., & Moseley, C. (2012). A three-tier diagnostic test to assess pre-service teachers' misconceptions about global warming, greenhouse effect, ozone layer depletion, and acid rain. *International Journal of Science Education*, 34(11), 1667–1686.
- Bunawan, W. (2014). Pengembangan Instrumen Tes Diagnostik Pilihan Ganda Tiga Tingkat Untuk Mengakses Kemampuan Mahasiswa Calon Guru Fisika. *EDUSAINS*, 6(2), 137–144.
- Caleon, I., & Subramaniam, R. (2010). Development and application of a three-tier diagnostic test to assess secondary students' understanding of waves. *International Journal of Science Education*, 32(7), 939–961.
- Creswell, J. W. (2012). *Research design pendekatan kualitatif, kuantitatif, dan mixed, Pustaka Pelajar*. Yogyakarta.
- Djanuarsih, E. (2012). Validitas dan reliabilitas butir soal. *E-Jurnal Dinas Pendidikan Kota Surabaya*, 1(1), 1–12.
- Embretson, S. E., & Reise, P. E. (2000). Item response theory for psychologists. Mahwah (NJ): Lawrence Earlbaum Associates. *Inc.[Google Scholar]*.
- Irwan, A. (2015). Efektifitas Penggunaan Model Pembelajaran Contextual Teaching And Learning (CTL) Dalam Meningkatkan Pemahaman Konsep Dan Keterampilan Praktikum Fisika. *JPF (Jurnal Pendidikan*

Fisika) Universitas Islam Negeri Alauddin Makassar, 3(1), 5–8.

- Maulini, S., Kurniawan, Y., & Mulyani, R. (2016). The Three Tier-Test untuk Mengungkap Kuantitas Siswa Yang Miskonsepsi Pada Konsep Gaya Pegas. *JIPF (Jurnal Ilmu Pendidikan Fisika), 1(2)*, 42–44.
- Natalia, D., Handhika, J., & Huriawati, F. (2017). Pengembangan Instrumen Tes Diagnosis Konsep IPA Fisika. *Momentum: Physics Education Journal, 103–110.*
- Peşman, H., & Eryılmaz, A. (2010). Development of a three-tier test to assess misconceptions about simple electric circuits. *The Journal of Educational Research, 103(3)*, 208–222.
- Popham, W. J. (2010). *Everything school leaders need to know about assessment.* Corwin Press.
- Prawira, W. Y. (2018). Analisis Miskonsepsi Mahasiswa Calon Guru Fisika Pada Konsep Kinematika Partikel Menggunakan Tes Diagnostik Three Tier Dan Wawancara Klinis. *Journal Fakultas Sains Dan Teknologi. UIN Walisongo, 1.*
- Purwanto, 2011. (n.d.). *Evaluasi Hasil Belajar.* Pustaka Pelajar.
- Qasem, M. A. N. (2013). A comparative study of classical theory (CT) and item response theory (IRT) in relation to various approaches of evaluating the validity and reliability of research tools. *IOSR Journal of Research & Method in Education (IOSR-JRME), 3(5)*, 77–81.
- Sudjana, N. (2009). Penilaian Hasil Proses Belajar Mengajar Remaja. *Rosdakarya: Bandung.*
- Uno, H. B., & Koni, S. (2013). *Assessment pembelajaran.*