

Pengembangan Instrumen Penilaian Model Two-Tier Multiple Choice Question (TTMCQ) pada Materi Pancadharma

Mirrah Megha Singamurti

STAB Negeri Raden Wijaya Wonogiri Jawa Tengah, Indonesia
Mirrahmegha99@gmail.com

Abstract

This study aimed to develop a Two Tier Multiple Choice Question Assessment instrument that is suitable for use in Pancadharma material and to find out the teachers and students' responses to the instrument developed. This study was conducted using the R&D method applying the steps from Borg and Gall. The locations of the study were at the State Junior High Schools 1 and 2 Keling in Jepara Regency. Data collection used questionnaires, test, interviews and documentation. The results of this study exhibited that the TTMCQ assessment instrument developed for the Pancadharma material had 92.22% of content validation; the reliability scores were 0.754 for first tier and 0.826 for two tier; the power of distinguishing questions was categorized into a good category with the percentage of 66.67% for the main questions, and 46.66% for the reason questions. The difficulty level test indicated that 40% of the multiple choice questions were easy and 53.33% of them were medium, and 60% of the reason questions were easy, and 33.33% of them were medium. The teachers and students' responses to the TTMCQ instrument developed in the form of multiple-choice multiple-level questions were more complicated than those of regular multiple-choice questions; the form of the questions was good and neatly organized; and the form of the questions was more about children's responsibility in terms of thinking and solving problems. Thus, the TTMCQ instrument that had been developed was included in the appropriate category to be used as an assessment instrument of the Pancadharma material.

Keywords: Assessment Instrument, Buddhist Education, Development Research, Two-Tier Multiple Choice Question, Panca Dharma.

Abstrak

Penelitian ini bertujuan untuk mengembangkan instrumen penilaian Two-Tier Multiple Choice Question (TTMCQ) yang layak digunakan pada materi Pancadharma dan mengetahui tanggapan guru dan siswa terhadap instrumen yang dikembangkan. Penelitian ini dilakukan menggunakan metode R&D dengan langkah-langkah dari Borg and Gall. Lokasi penelitian dilakukan di SMP Negeri 1 Keling dan 2 Keling di Kabupaten Jepara. Pengumpulan data menggunakan angket, tes, wawancara dan dokumentasi. Hasil penelitian ini menunjukkan bahwa instrumen penilaian TTMCQ materi Pancadharma yang dikembangkan memiliki validasi isi sebesar 92,22%; reliabilitas untuk first tier sebesar 0,754 dan two tier sebesar 0,826; daya pembeda yang baik pada soal utama sebesar 66,67% dan soal alasan sebesar 46,66%. Uji tingkat kesukaran soal pilihan ganda menunjukkan 40% soal mudah dan 53,33% soal sedang. Uji tingkat kesukaran soal alasan menunjukkan 60% soal dalam kategori mudah dan 33,33% soal dalam kategori sedang. Tanggapan guru dan siswa terhadap instrumen TTMCQ yang dikembangkan soal dalam bentuk pilihan ganda bertingkat lebih rumit dibandingkan dengan pilihan ganda biasa, bentuk soal sudah bagus dan tertata rapi, bentuk soal lebih menantang kemampuan anak dalam berpikir dan menyelesaikan soal. Dengan demikian instrumen TTMCQ yang telah dikembangkan masuk dalam kategori layak untuk digunakan sebagai instrumen penilaian pada materi Pancadharma.

Kata kunci: Instrumen Penilaian, Panca dharma, Pendidikan Agama Buddha, Penelitian Pengembangan, Two-Tier Multiple Choice Question.

Permalink/DOI: <https://doi.org/10.18326/infsl3.v14i1.97-120>

Pendahuluan

Pendidikan adalah suatu usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia serta keterampilan yang diperlukan dirinya, masyarakat, bangsa dan negara. Fungsi dari pendidikan nasional adalah mengembangkan kemampuan dan membentuk watak serta peradaban bangsa yang bermartabat dalam rangka mencerdaskan kehidupan bangsa, selain itu, pendidikan nasional bertujuan untuk berkembangnya potensi peserta didik agar menjadi manusia yang beriman dan bertakwa kepada Tuhan Yang Maha Esa, berakhlak mulia, sehat, berilmu, cakap, kreatif, mandiri dan menjadi warga negara yang demokratis serta bertanggung jawab (UU RI nomor 20 tahun 2003 pasal 1 dan 3 tentang sistem pendidikan nasional).

Proses pendidikan yang berkualitas akan menghasilkan pendidikan yang sesuai dengan fungsi dan tujuannya. Proses pendidikan merupakan kegiatan sosial atau pergaulan antara pendidik dan peserta didik dengan menggunakan isi atau materi pendidikan, metode dan alat pendidikan tertentu yang berlangsung dalam suatu lingkungan untuk pencapaian tujuan pendidikan yang telah ditetapkan. Kualitas pendidikan sangat ditentukan oleh kemampuan satuan pendidikan dalam mengelola proses pembelajaran. Salah satu kualitas dari pelaksanaan tugas keguruan, dapat dilihat dari hasil proses yang dilaksanakannya. Komponen mengajar maupun mendidik, melatih dan membimbing, tidak terpisahkan dengan tiga komponen penting sebagai hasil dari pendidikan, berupa pengetahuan, keterampilan dan sikap. Ketiga komponen hasil belajar tersebut dapat diketahui melalui komponen menilai dan mengevaluasi. Kesemua komponen tersebut terkait dengan kompetensi guru, dalam pelaksanaan tugasnya.

Berdasarkan hasil orientasi yang dilaksanakan di SMP Negeri 1 dan 2 Keling Kabupaten Jepara, khususnya untuk pendidikan agama Buddha, masih terdapat permasalahan guru dalam melaksanakan tugasnya. Beberapa diantaranya terkait dengan pemahaman dengan keterbatasan informasi dan pengembangan diri guru, terbatasnya

kegiatan pengembangan diri, sehingga berakibat kesulitan dan keterbatasan dalam pengembangan bahan ajar, serta aplikasi model penilaian dari materi bahan ajar, yang tidak berkembang. Model penilaian pendidikan Agama Buddha masih terbatas pada tes tertulis, tes lisan dan penugasan atau proyek, guru belum mengembangkan instrumen pengembangan yang lain, sehingga hasil penilaian belum dapat digunakan untuk mengungkap pengetahuan yang dimiliki peserta didik secara mendalam tanpa takut adanya subjektivitas penilaian dan juga dapat mengurangi tindakan curang siswa dalam menjawab.

Dalam pendidikan Agama Buddha belum banyak dilakukan pengembangan penilaian pembelajaran. Hal ini dapat diketahui dari evaluasi pembelajaran yang dipergunakan guru masih terbatas pada pilihan ganda dan essay, guru lebih cenderung menggunakan satu jenis soal saja, terkadang pilihan ganda saja atau essay saja, hal ini juga terlihat dari evaluasi yang dipergunakan guru dalam materi Panca Dharma. Penggunaan instrumen penilaian dalam bentuk pilihan ganda hanya dapat menilai konten pengetahuan tanpa mempertimbangkan alasan dibalik pilihan tersebut. Sebagai perbaikan dari hal ini perlu dikembangkan suatu instrumen pilihan ganda yang mencakup respon dan alternatif konsepsi siswa. Siswa diharuskan untuk membenarkan pilihan jawaban mereka dengan memberikan alasan. Penggunaan alasan ketika menjawab item tes pilihan ganda menjadi cara yang sensitif dan efektif untuk menilai pembelajaran yang bermakna. Penilaian semacam ini dapat dilakukan melalui jenis soal pilihan ganda dua tingkat, Two Tier Multiple Choice Question (TTMCQ). Hal positif temuan yang terkait dengan alasan siswa menyebabkan perkembangan tes diagnostik pilihan ganda dua tingkat TTMCQ bertujuan khusus untuk mengidentifikasi alternatif konsepsi siswa dalam konten yang jelas dan terbatas.

Selain itu, lingkungan masyarakat di Kecamatan Keling Kabupaten Jepara yang banyak terdapat umat Buddha sangat mengharapkan pendidikan Agama Buddha yang diajarkan di sekolah selaras dengan kebutuhan di masyarakat. Siswa diharapkan dapat mengaplikasikan hasil pembelajarannya di sekolah dalam kehidupan bermasyarakat. Materi pendidikan Agama Buddha yang relevan diantaranya adalah Panca dharma yang menekankan

pemahaman dan aplikasi siswa kepada lingkungan masyarakat. Salah satu yang dapat dinilai yaitu perilaku siswa dalam keseharian sebagai bukti pemahaman dari konsep dan aplikatif dalam materi pembelajaran terutama Pendidikan agama Buddha yang diterima di sekolah.

Berdasarkan uraian diatas perlu adanya pengembangan instrumen penilaian *Two-Tier Multiple Choice Question* (TTMCQ) pada materi Pancadharma untuk tingkat Sekolah Menengah Pertama di Jepara tepatnya di SMP Negeri 1 Keling dan SMP Negeri 2 Keling Kabupaten Jepara.

Kajian Teori

Penilaian Pembelajaran

Penilaian merupakan bagian yang penting dalam proses pembelajaran. Dengan melakukan penilaian, pendidik sebagai pengelola kegiatan pembelajaran dapat mengetahui kemampuan yang dimiliki peserta didik, ketepatan metode mengajar yang digunakan dan keberhasilan peserta didik dalam meraih kompetensi yang ditetapkan. Dalam pelaksanaan menilai dan mengevaluasi, didalam komponen penilaian terkait dengan aspek penilaian pengetahuan yang merupakan aspek penting dan menjadi tolok ukur kemampuan siswa dalam memahami materi atau kompetensi yang diajarkan. Penilaian dalam proses pembelajaran digunakan sebagai alat untuk mendeteksi kesulitan belajar (*assessment as learning*), penilaian sebagai proses pembelajaran (*assessment for learning*), dan penilaian sebagai alat untuk mengukur pencapaian dalam proses pembelajaran (*assessment of learning*). Melalui penilaian tersebut diharapkan peserta didik dapat menguasai kompetensi yang diharapkan. Untuk itu, digunakan teknik penilaian yang bervariasi sesuai dengan kompetensi yang akan dinilai, yaitu testulis, lisan, dan penugasan.

Selain itu, Sudijono (2005) menyatakan identifikasi terhadap seluruh (instrumen) item tes hasil belajar perlu dilakukan dengan harapan akan menghasilkan berbagai informasi berharga yang pada dasarnya merupakan umpan balik guna melakukan perbaikan, pembenahan, penyempurnaan kembali terhadap butir-butir item yang dikeluarkan dalam tes hasil belajar.

Two-tier Multiple Choice

Adodo (2013) menjelaskan bahwa Two-tier Multiple Choice (TTMC) adalah bentuk pertanyaan yang lebih canggih dari pertanyaan pilihan ganda. Tingkat pertama menyerupai pilihan ganda tradisional, yang biasanya berkaitan dengan pernyataan pengetahuan. Tingkat kedua menyerupai format dari soal pilihan ganda tradisional tetapi bertujuan untuk mendorong pemikiran dan penalaran. Sementara itu, banyak guru tidak dapat menyiapkan instrumen Two-Tier Multiple Choice (TTMC) dengan versi mereka sendiri karena kurangnya waktu atau kurangnya pengetahuan tentang pengembangan tes jenis ini.

Research and Development (R&D)

Penelitian dan pengembangan (R&D) adalah proses yang digunakan untuk mengembangkan dan memvalidasi produk pendidikan. Langkah dalam proses ini biasanya merujuk pada siklus R&D yang terdiri dari mempelajari temuan penelitian untuk kemudian mengembangkan produk, mengembangkan produk berdasarkan temuan penelitian, uji lapangan dimana produk itu akan digunakan nantinya, dan merevisi produk untuk memperbaiki kekurangan yang ditemukan pada uji lapangan. Dalam siklus R&D ini akan berulang sampai produk yang dikembangkan sesuai dengan tujuan yang telah ditetapkan (Borg dan Gall, 1983: 624).

Langkah utama dalam siklus R&D (Borg dan Gall, 1983:624-625) yang digunakan untuk mengembangkan produk adalah perencanaan (*planning*), pengembangan bentuk awal produk (*developing of the preliminary form of the product*), uji lapangan awal dan revisi (*preliminary field test and product revision*), uji lapangan utama dan Revisi (*Main field test and product revision*), uji pelaksanaan lapangan dan revisi (*Operational field test and final product revision*), diseminasi dan implementasi (*dissemination and implementation*).

Perencanaan merupakan aspek paling penting dari penelitian yang berbasis produk pendidikan yang merupakan pernyataan tujuan spesifik yang akan dicapai dari produk yang akan dikembangkan. Pengembangan bentuk awal produk dapat didasarkan pada hasil angket dan wawancara. Langkah ini merupakan suatu prinsip

yang penting yang harus diamati dalam pengembangan produk bentuk awal sehingga memungkinkan mendapatkan banyak masukan dari uji lapangan. Uji lapangan awal dan revisi merupakan langkah yang dilakukan untuk mendapatkan evaluasi kuantitatif awal dari produk pendidikan yang baru sehingga hasilnya bisa digunakan sebagai acuan revisi produk. Uji lapangan utama dan revisi dilakukan untuk menentukan apakah produk yang sedang dikembangkan peformanya memenuhi tujuannya. Hasil dari uji ini digunakan sebagai landasan untuk merevisi produk. Uji pelaksanaan lapangan dan revisi dilakukan dengan tujuan untuk menentukan produk pendidikan telah benar-benar siap untuk digunakan di sekolah tanpa kehadiran pengembang dan stafnya. Hasil dari uji ini digunakan sebagai saran dan masukan untuk penyempurnaan produk yang dikembangkan. Diseminasi dan Implementasi merupakan cara yang efektif untuk menjustifikasi dengan mendemostrasikan produk hasil penelitian dan pengembangan kepada khalayak.

Metode Penelitian

Penelitian ini dilaksanakan menggunakan metode penelitian dan pengembangan (R&D). Sampel adalah bagian dari populasi. Keberadaan sampel mewakili populasi. Dalam penelitian ini menggunakan teknik sampling purposive sampling yaitu pemilihan sampel dengan pertimbangan tertentu (Musfiqon, 2012:96). Penelitian mempertimbangkan kesesuaian masalah dengan individu yang dipilih. Pemilihan individu yang disajikan sampel ini didasarkan pada pengetahuan dan informasi yang diperoleh peneliti tentang individu yang ditunjuk tersebut. Pertimbangan yang dimaksud adalah pengambilan sampel yang mewakili informasinya bukan populasinya. Informan dalam penelitian ini terdiri dari ahli atau pakar, peserta didik serta guru. Ahli atau pakar dipilih berdasarkan pada kompetensi dalam bidangnya dan memiliki kredibilitas untuk menjadi sumber data. Sekolah yang dipilih merupakan sekolah dengan kriteria baik berdasarkan input peserta didik maupun rentang nilai yang di capai berdasarkan mata pelajaran pendidikan agama buddha dengan jumlah 2 sekolah.

Populasi dan Sampel yang akan digunakan dalam penelitian ini adalah siswa kelas VII dan VIII SMP Negeri 1 Keling dan SMP Negeri 2 Keling yang beragama Buddha, selain siswa juga melibat-

kan guru pendidikan Agama Buddha dan mahasiswa STAB Negeri Raden Wijaya Wonogiri. Skala kecil akan dilaksanakan dengan melibatkan 2 orang guru pendidikan agama buddha yang berasal dari SMP Negeri 1 Keling dan SMP Negeri 2 Keling serta 3 Mahasiswa dari STAB Negeri Raden Wijaya Wonogiri yang melaksanakan PPL di SMP Negeri 1 Keling dan SMP Negeri 2 Keling. Pelaksanaan uji coba skala kecil dilakukan saat di SMP Negeri 1 Keling dan SMP Negeri 2 Keling. Pelaksanaan uji coba kecil untuk mengetahui bagaimana kualitas soal, kedalaman soal yang dibahas serta kesesuaian soal yang dibuat untuk kategori siswa tingkat SMP.

Skala Menengah akan dilaksanakan dengan melibatkan siswa kelas VIII sebanyak 10 yang terbagi dari 5 siswa dari SMP Negeri 1 Keling dan 5 siswa dari SMP Negeri 2 Keling, digunakan siswa kelas VIII karena siswa kelas VIII yang sudah pernah memperoleh materi Panca Dharma. Subjek yang dipergunakan dalam uji coba skala menengah ini adalah siswa yang beragama buddha di sekolah tersebut dan kelas VIII, pengambilan sampel untuk siswa uji coba skala menengah dilakukan dengan acak. Dilakukan uji coba skala menengah dengan tujuan untuk mengetahui kelayakan soal, kedalaman soal dan kesesuaian soal yang dipergunakan dalam penilaian ini.

Skala Meluas akan dilaksanakan dengan melibatkan siswa kelas VII sebanyak 20 yang terbagi dari 10 siswa dari SMP Negeri 1 keling dan 10 siswa dari SMP Negeri 2 Keling. Pertimbangan menggunakan kelas VII karena materi Panca Dharma disampaikan di kelas VII. Dipergunakan siswa kelas VII ini disesuaikan dengan materi ajar yang dipergunakan yaitu kelas VII dan menggunakan siswa berjumlah 10 dari SMP Negeri 1 Keling dan 10 siswa dari SMP Negeri 2 Keling, pengambilan sampel ini dilakukan secara acak. Uji skala meluas ini untuk mengetahui kelayakan serta efektifitas soal yang dibuat.

Teknik dan pengumpulan data yang digunakan dalam penelitian ini antara lain: 1) angket, 2) lembar wawancara, 3) tes. Sedangkan untuk teknik analisis data menggunakan validitas isi dimana dalam penelitian ini menggunakan validasi isi dari ahli yang dianalisis dengan menggunakan formula Aiken. Uji Validasi yang digunakan adalah validitas isi, dimana nilai Validitas berkisar antara 0 - 1 dan

suatu butir soal dikatakan valid secara isi apabila memiliki harga Validitas lebih dari 0,7 yang diperoleh dari penilaian validasi ahli.

Pengujian reliabilitas menggunakan *Alpha Cronbach* menggunakan bantuan aplikasi SPSS 18, dimana instrument dapat dikatakan reliabel apabila hasil analisis reliabilitas diperoleh nilai lebih dari 0,7. Apabila nilai yang diperoleh melalui analisis SPSS 18 diperoleh nilai reliabilitas kurang dari 0,7 maka instrument yang dibuat dinyatakan tidak reliabel, secara otomatis apabila dinyatakan tidak reliabel instrument harus dilakukan perbaikan dan penilaian ulang sampai diperoleh analisis *Alpha Cronbach* lebih dari 0,7.

Selain dilakukan uji validitas isi dengan menggunakan aiken dan reliabilitas instrument dengan menggunakan *Alpha Cronbach*, dilakukan pula pengujian daya beda soal, dimana analisis daya beda item yang digunakan untuk melihat besar kecilnya angka indeks diskriminasi item, dalam penelitian ini menggunakan indeks diskriminasi (daya pembeda) dengan menggunakan dasar perhitungan pada pembagian testee kedalam dua kelompok, kelompok atas dan kelompok bawah.

Hasil dan Pembahasan

Penelitian ini dimulai dengan melaksanakan uji pendahuluan (*preliminary study*) di sekolah. tahap uji pendahuluan dimaksudkan untuk menganalisis kebutuhan (*need assessment*) di lapangan, sehingga produk yang dikembangkan benar-benar berasal dari kebutuhan yang ada di lapangan. Uji pendahuluan dilaksanakan dengan melakukan wawancara dengan guru dan siswa mata pelajaran Pendidikan Agama Buddha. Hasil dari uji pendahuluan yang dilakukan diperoleh hasil bahwa guru belum banyak melakukan inovasi untuk pembelajaran baik dari perencanaan, pelaksanaan dan evaluasi. Masih terbatas untuk pelatihan dalam pengembangan dalam pembelajaran sehingga masih sangat sedikit pengembangan yang dilakukan oleh guru.

Guru juga menjelaskan bahwa evaluasi pembelajaran yang dilakukan belum dilakukan pengembangan dalam hal ini hanya terbatas pada soal essay dan pilihan ganda. Belum adanya pelatihan ataupun contoh soal yang lain membuat guru kesulitan dalam melakukan pengembangan dalam pembelajaran. Hasil wawancara

dengan siswa diperoleh data bahwa di SMP Negeri 1 Keling maupun di SMP Negeri 2 Keling untuk evaluasi siswa baik ulangan hari, UTS, maupun UAS hanya terbatas pada soal essay dan pilihan ganda saja. Belum pernah dilakukan evaluasi menggunakan model soal yang lain, apabila mengerjakan soal yang ada di buku pendamping kadang menggunakan jenis soal yang lain seperti teka teki silang.

Pengembangan Produk

Pengembangan produk dalam penelitian ini yaitu pengembangan instrument penilaian menggunakan model *Two Tier Multiple Choice Question (TTMCQ)* pada materi Pancadharmas. Materi ini terdapat pada bagian Standar Kompetensi Sila: 3. Konstruksi sikap umat Buddha untuk menjadi manusia susila sesuai Pancadharmas, Kompetensi Dasar pada bagian 3.2. menguraikan Pancadharmas dan 3.3 mendeskripsikan Pancadharmas dalam kehidupan sehari-hari.

Materi yang terkandung didalamnya yaitu membahas tentang *Metta karuna, Samma ajiva, Kamasavara, Sacca, Sati sampajana* serta penerapannya dalam kehidupan sehari-hari. Dalam penelitian ini dilakukan pengembangan instrument pembelajaran, dimana instrument pembelajaran yang dikembangkan yaitu alat evaluasi yang dipakai, dalam penelitian ini mengembangkan model penilaian *Two Tier Multiple Choice Question* dimana dalam penilaian tersebut terdapat dua jenis soal yaitu soal utama dan soal alasan, soal utama berisikan materi pokok yang berkaitan dengan indikator yang telah dibuat, sedangkan soal alasan berisikan alasan yang mendukung jawaban soal utama, dimana alasan ini harus berkaitan dengan setiap soal utama yang ada. Desain awal penilaian yang dibuat berjumlah 15 soal tipe *TTMCQ*

Penyusunan Draft Produk

Hasil pencarian informasi melalui hasil studi pustaka, hasil wawancara dengan guru dan siswa dibuat produk penelitian pengembangan instrument penilaian menggunakan model *Two Tier Multiple Choice Question* dengan menggunakan materi Pancadharmas. Evaluasi yang digunakan yaitu pada dasarnya masih menggunakan pilihan ganda tetapi merupakan pilihan ganda bertingkat, dimana soal tersebut terdapat soal utama dan alasan. Dalam evaluasi ini dapat diketahui siswa memahami konsep atau tidak, karena ti-

dak hanya soal pilihan ganda saja tetapi terdapat alasan yang dapat memperkuat jawaban soal utama. Jika siswa memiliki pengetahuan yang kuat terhadap materi maka dapat diketahui tidak hanya merupakan hafalan saja. Pengetahuan yang dimiliki terkait materi juga akan dapat diketahui, seberapa paham siswa dalam mempelajari materi yang dijelaskan guru pun akan sangat terlihat.

Sebelum menyusun produk perlu diperhatikan beberapa hal diantaranya pemeriksaan pendahuluan (*preliminary investigation*) dan penyesuaian teori (*Theoretical embedding*). Tahap awal pemeriksaan pendahuluan ini merupakan tahap awal dilakukan penelitian. Pada tahap ini dilakukan analisis kebutuhan (*need Assesment*) dilapangan guna memperoleh informasi yang benar-benar dibutuhkan di sekolah. Tahap pemeriksaan pendahuluan dalam penelitian ini dilakukan dengan wawancara yang melibatkan siswa dan guru di SMP Negeri 1 Keling dan SMP Negeri 2 Keling Kabupaten Jepara.

Hasil dari tahap pemeriksaan pendahuluan penelitian guru hanya menggunakan soal evaluasi berupa soal model essay dan pilihan ganda. Namun soal pilihan ganda masih kurang efektif untuk mengetahui kemampuan pemahaman yang dimiliki oleh siswa, sedangkan soal essay atau uraian membutuhkan waktu yang lama dalam pelaksanaan dan pengoreksiannya. Berdasarkan hal tersebut dalam penelitian ini instrument yang dipilih *Two-Tier Multiple Choice Question*. Bentuk soal ini merupakan jenis soal dengan pilihan jawaban utama disertai dengan alasan.

Tahap penyesuaian teoritis (*theoretical embedding*) bertujuan untuk meningkatkan pemahaman secara teoritis tentang produk pendidikan, terjaminnya bahwa produk pendidikan yang dikembangkan memiliki landasan teori yang kukuh, sehingga secara ilmiah dapat dipertanggung jawabkan (Sanjaya, 2013). Pada tahap ini dilakukan pengakajian literature tentang pengembangan penilaian pembelajaran dengan *TTMCQ* pada materi Pancadharama untuk siswa tingkat SMP. Langkah yang dilakukan yaitu pertama merencanakan tujuan tes, merumuskan lingkup materi tes, menentukan kompetensi yang akan di uji, menyusun kisi-kisi, pengembangan draf instrumen.

Langkah pertama merencanakan tujuan tes yaitu sebagai intru-

men penilaian formatif, untuk memperoleh masukan tentang keberhasilan pelaksanaan pembelajaran, dimana masukan tersebut akan berguna untuk memperbaiki strategi dalam pembelajaran. Langkah kedua merumuskan lingkup materi tes tentang pancadharmas dimana materi tersebut erat kaitannya dengan materi sila terutama Pancasila buddhis yang ada, dalam materi pancadharama harus berisikan tentang *Metta Karuna, Samma Ajiva, Kamasavara, Sacca, Sati Sampajanna* dimana kesemua materi tersebut berkaitan dengan kehidupan sehari-hari. Langkah ketiga yaitu menentukan kompetensi yang akan diuji atau indikator yang akan diuji disesuaikan dengan indikator yang akan dicapai. Sekaligus penyusunan kisi-kisi dengan materi yang ada dengan tipe soal, karena tipe soal utama dengan tipe soal alasan berlainan. Langkah terakhir yaitu pengembangan draf instrument, dimana soal yang dibuat disesuaikan dengan indikator soal yang telah dibuat sebelumnya. Dimana penulisan soal adalah penjabaran indikator menjadi pertanyaan-pertanyaan yang dikarakteristiknya sesuai dengan pedoman kisi-kisi. Penskoran model soal *two-tier test* yang akan disajikan dalam tabel 1

Tabel 1. Skoring Soal *Two Tier Multiple Choice Question*

| No | Aspek Penilaian | Skor |
|----|----------------------------------------------------------|------|
| 1 | Tidak menjawab (TJ) atau jawaban salah-alasan salah (SS) | 0 |
| 2 | Jawaban salah-alasan benar (SB) | 1 |
| 3 | Jawaban benar-alasan salah (BS) | 2 |
| 4 | Jawaban benar-alasan benar (BB) | 3 |

Dengan prosedur penskoran tersebut akan digunakan untuk menganalisis kemampuan siswa. Selain itu akan dapat dideteksi letak kelemahan siswa sebagai fungsi diagnostik

Uji Coba Draft Produk

Pada tahap ini diawali dengan uji validitas draft produk. Uji Validitas dalam penelitian ini menggunakan validitas isi. Validitas isi berkaitan dengan sejauh mana keseluruhan materi atau bahan yang diukur. Validitas dilakukan oleh 3 orang ahli dalam evaluasi. Dalam penelitian ini, validator butir soal dilakukan oleh Suharno, M.Pd.B selaku dosen agama Buddha mengajar Kitab Suci Agama Buddha Suttapitaka, Ari Mariyono, M.Pd sebagai dosen Pokok-Pokok Dasar

Agama Buddha, Kiswati, S.Ag sebagai guru pengajar agama Buddha di SMP Negeri 1 Keling dan SMP Negeri 2 Keling Kabupaten Jepara serta 5 mahasiswa dari Sekolah Tinggi Agama Buddha Negeri Raden Wijaya Wonogiri Semester 6 yang sudah menempuh PPL (Praktik Mengajar).

Hasil validasi yang telah dilakukan digunakan sebagai dasar revisi penilaian pembelajaran. Validasi yang dilakukan menggunakan validasi Aiken. Validasi isi dilakukan oleh 2 orang ahli perwakilan dari dosen agama Buddha dan guru agama Buddha serta 5 mahasiswa dari STAB Negeri Raden Wijaya Wonogiri Jawa Tengah dengan perolehan hasil validasi 0,9222 yang menyatakan dapat dilakukan analisis. Dari analisis hasil validasi draf produk penilaian tersebut diperoleh kesimpulan bahwa penilaian TTMCQ dengan materi Pancadharmas yang dibuat layak untuk digunakan dan dapat dilanjutkan untuk penelitian.

Pengujian Produk

Produk yang telah dikembangkan dilakukan beberapa kali uji coba. Uji coba dilakukan sebanyak tiga tahap, yaitu uji coba skala terbatas, uji coba skala menengah dan uji coba skala luas. Tiga uji coba yang telah dilakukan diperoleh hasil sebagai berikut.

Pertama, uji coba skala terbatas bertujuan untuk mengetahui keterbacaan penilaian *Two Tier Multiple Choice Question*. Sampel untuk keterbacaan di berikan kepada guru dan calon pendidikan disini peneliti melibatkan 5 mahasiswa STAB Negeri Raden Wijaya jurusan Dharmacarya program studi Pendidikan Keagamaan Buddha untuk menilai instrumen penilaian TTMCQ dengan materi Pancadharmas. Pengujian ini dilakukan dengan tujuan supaya dapat memberikan masukan terkait dengan materi sekaligus mengetahui kesesuaian penilaian dengan aturan yang ada. 5 mahasiswa dipilih secara acak, hasil uji coba skala terbatas ini digunakan sebagai dasar untuk revisi awal. Angket keterbacaan penilaian pada uji coba skala terbatas yang diberikan pada guru dan siswa diperoleh hasil sebagai

Tabel 2. Hasil Uji Coba Skala Terbatas Penilaian TTMCQ

| Instrumen | Sebelum Revisi | Sesudah Revisi |
|------------------|------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|
| Penilaian TTMCQ | Kalimat yang digunakan pada penilain kurang dapat dimengerti siswa dan masih menimbulkan makna ganda | Kalimat yang digunakan pada penilaian TTMCQ dibuat menjadi sederhana sehingga siswa mudah memahaminya |
| | Masih terdapat soal yang belum terdapat keterkaitan antara soal utama dengan soal alasan | Soal dibuat sederhana dan tidak berbelit-belit tetapi memiliki benang merah antara soal utama dengan soal alasan |
| | Kalimat pada setiap pilihan jawaban sama antar soal sehingga membuat kerancuan makna | Kalimat pada pilihan jawaban dirubah dengan kalimat yang memiliki makna sama |

Kedua, uji coba skala menengah bertujuan untuk mengetahui karakteristik produk penilaian TTMCQ yang dibuat. Sampel untuk uji coba skala menengah adalah siswa kelas 8 baik di SMP Negeri 1 Keling maupun SMP Negeri 2 Keling. Hasil dari uji skala menengah ini digunakan sebagai dasar untuk merevisi penilaian TTMCQ yang telah dibuat, yang nantinya akan dipergunakan untuk analisis butir soal dan uji keberlanjutan. Pada uji coba skala menengah penilaian yang dikembangkan sudah dapat diaplikasikan untuk uji coba skala luas. Dalam hal ini perbaikan penilaian TTMCQ disajikan dalam tabel 3 berikut ini.

Tabel 3. Hasil Uji Coba Skala Menengah Produk Penilaian TTMCQ

| Produk | Sebelum Revisi | Sesudah Revisi |
|------------------|-------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|
| Penilaian TT-MCQ | Waktu yang digunakan untuk mengerjakan soal terlalu singkat | Waktu pengujian selanjutnya diberikan tambahan waktu sehingga waktu yang digunakan untuk mengerjakan tercukupi |

Pada tahap uji coba skala menengah, dilakukan pengujian mengenai reliabilitas, tingkat kesukaan dan daya pembeda menggunakan software Itean. Soal yang diujikan pada uji coba skala menengah ber-

jumlah 15 soal. Adapun hasil pengujian soal TTMCQ pada uji coba skala menengah sebagai berikut.

Reliabilitas

Pengujian reliabilitas dalam penelitian ini menggunakan alpha yang tertera pada iteman. Adapun hasil pengujian reliabilitas soal two tier test secara keseluruhan pada uji coba skala menengah disajikan pada tabel 4 berikut ini.

Tabel 4. Hasil Perhitungan Reliabilitas Skala Menengah

| Uji Coba Skala Menengah | Nilai |
|-------------------------|-------|
| <i>First Tier</i> | 0,754 |
| <i>Second Tier</i> | 0,826 |

Pengujian reliabilitas soal hanya dilakukan pada pengujian skala menengah. Hal ini dengan tujuan dari masing-masing uji coba, dimana uji coba skala menengah digunakan untuk mengetahui karakteristik produk yang dibuat sehingga perhitungan karakteristik soal hanya dilakukan pada hasil pada uji coba skala menengah. Berdasarkan hasil pengujian reliabilitas soal TTMCQ tersebut, reliabilitas yang dihasilkan yaitu $r > 0,700$. Hal ini berarti bahwa soal TTMCQ yang diujikan dapat dikatakan reliabel sehingga memiliki tingkat keajegan atau konsistensi yang baik.

Tingkat kesukaran

Pengujian tingkat kesukaran menggunakan software Iteman. Tingkat kesukaran soal menunjukkan sukar atau tidaknya butir soal, dimana tingkat kesukuran dilihat dari peluang siswa menjawab benar pada tingkat kemampuan tertentu. Adapun persentase tingkat kesukaran soal dan hasil pengujian tingkat kesukaran soal TTMCQ pada tabel 5 dan 6 berikut ini.

Tabel 5. Persentase Tingkat Kesukaran Soal

| Tingkat Kesukaran | Persentase (%) | | |
|--------------------|----------------|----------------|-------|
| | Mudah | Cukup (Sedang) | Sukar |
| <i>First Tier</i> | 40 | 53,33 | 6,67 |
| <i>Second Tier</i> | 60 | 33,33 | 6,67 |

Tabel. 6. Hasil Perhitungan Tingkat Kesukaran Soal

| | No Soal | Tingkat Kesukaran | |
|-----|-------------|-------------------|------------|
| | | Indeks | Keterangan |
| | First tier | 0,571 | C |
| 1. | Second tier | 0,071 | S |
| | First tier | 0,500 | C |
| 2. | Second tier | 0,429 | C |
| | First tier | 0,857 | M |
| 3. | Second tier | 0,929 | M |
| | First tier | 0,071 | S |
| 4. | Second tier | 0,786 | M |
| | First tier | 0,643 | C |
| 5. | Second tier | 0,857 | M |
| | First tier | 0,786 | M |
| 6. | Second tier | 0,786 | M |
| | First tier | 0,929 | M |
| 7. | Second tier | 0,714 | M |
| | First tier | 0,643 | C |
| 8. | Second tier | 0,429 | C |
| | First tier | 0,786 | M |
| 9. | Second tier | 0,429 | C |
| | First tier | 0,571 | C |
| 10. | Second tier | 0,857 | M |
| | First tier | 0,429 | C |
| 11. | Second tier | 0,643 | C |
| | First tier | 0,857 | M |
| 12. | Second tier | 0,857 | M |
| | First tier | 0,714 | M |
| 13. | Second tier | 0,643 | C |
| | First tier | 0,357 | C |
| 14. | Second tier | 0,714 | M |
| | First tier | 0,643 | C |
| 15. | Second tier | 0,714 | M |

Keterangan:

M = Mudah

C = Cukup

S = Sukar

Daya Beda

Daya beda soal adalah kemampuan suatu butir soal untuk dapat membedakan siswa kelompok atas (siswa dengan skor tinggi) dengan kelompok bawah (siswa dengan skor rendah). Suatu butir soal dikatakan baik jika harga indeks bedanya bernilai positif, apabila indeks daya beda soal bernilai negatif, hal ini berarti bahwa jumlah siswa kelompok bawah lebih banyak menjawab benar dibandingkan dengan siswa kelompok atas yang menjawab benar pada butir soal tersebut. Adapun hasil pengujian daya beda soal pada indeks uji coba skala menengah dan persentasenya disajikan pada tabel 7 dan 8 berikut ini.

Tabel 7. Hasil Pengujian Daya Pembeda Soal

| No Soal | Daya Pembeda Soal | | |
|---------|-------------------|------------|----|
| | rpbis | Keterangan | |
| | First tier | 0,147 | K |
| 1. | Second tier | 0,338 | C |
| | First tier | 0,533 | B |
| 2. | Second tier | 0,504 | B |
| | First tier | 0,510 | B |
| 3. | Second tier | 0,804 | BS |
| | First tier | 0,148 | K |
| 4. | Second tier | 0,695 | B |
| | First tier | 0,521 | C |
| 5. | Second tier | 0,704 | BS |
| | First tier | 0,568 | B |
| 6. | Second tier | 0,644 | B |
| | First tier | 0,841 | BS |
| 7. | Second tier | 0,485 | B |
| | First tier | 0,231 | C |
| 8. | Second tier | 0,504 | B |

| No Soal | Daya Pembeda Soal | | |
|---------|-------------------|------------|----|
| | rpbis | Keterangan | |
| 9. | First tier | 0,568 | B |
| | Second tier | 0,334 | C |
| 10. | First tier | 0,475 | B |
| | Second tier | 0,583 | B |
| 11. | First tier | 0,555 | B |
| | Second tier | 0,627 | B |
| 12. | First tier | 0,643 | B |
| | Second tier | 0,824 | BS |
| 13. | First tier | 0,483 | B |
| | Second tier | 0,363 | C |
| 14. | First tier | 0,494 | B |
| | Second tier | 0,718 | BS |
| 15. | First tier | 0,618 | B |
| | Second tier | 0,299 | C |

Keterangan:

K = Kurang

C = Cukup

B = Baik

BS = Baik Sekali

Tabel 8. Persentase Daya Pembeda Soal

| Instrumen | Jumlah Item | Persentase Daya Beda Soal (%) | | | |
|--------------------|-------------|-------------------------------|-------|-------|-------------|
| | | Jelek | Cukup | Baik | Baik Sekali |
| <i>First Tier</i> | 15 | 13,33 | 13,33 | 66,67 | 6,67 |
| <i>Second Tier</i> | 15 | 0 | 26,67 | 46,66 | 26,67 |

Analisis Kualitas Soal

Analisis kualitas soal diperlukan untuk mengetahui kualitas soal yang dikembangkan. Analisis kualitas soal dilakukan berdasarkan pada hasil uji coba soal. Uji coba soal perlu dilakukan sebelum mengujikan

soal dalam tes yang sebenarnya, dimana hasil uji coba soal memberikan informasi atau data empirik tentang kebaikan soal yang dikembangkan (Mardapi, 2012). Berdasarkan hasil uji coba soal dapat diketahui: validitas, reliabilitas, daya pembeda dan tingkat kesukaran soal yang dikembangkan. Suatu soal dapat dikatakan sebagai soal yang baik apabila memenuhi kriteria validitas, reliabilitas, daya pembeda, tingkat kesukaran sebagai soal yang baik.

Uji Validasi yang digunakan adalah validitas isi, dimana nilai V berkisar antara 0 -1 dan suatu butir soal dikatakan valid secara isi apabila memiliki harga V yaitu 0,9222 untuk jumlah penilai (*raters*) 8 orang (Aiken, 1985). Berdasarkan uji validitas yang telah dilakukan dengan jumlah penilai (*raters*) 8 orang, soal *two-tier multiple choice question* yang dikembangkan dengan jumlah 15 butir, 1 butir soal dengan nilai validasi 0,83333, 1 butir soal dengan nilai 0,875, 8 butir soal dengan nilai validasi 0,91667, 5 butir soal dengan nilai validasi 0,95833, kesemua butir soal dinyatakan valid, tetapi dari soal tersebut terdapat masukan dari penilai sehingga dilakukan perbaikan.

Hasil Analisis Uji Coba Produk

Uji Coba Skala Terbatas

Uji coba skala terbatas dilakukan dengan mengujicobakan soal TT-MCQ materi pancadharma kepada 5 mahasiswa. Dimana dilakukan uji coba skala terbatas ini adalah untuk mengetahui keterbacaan soal, yaitu apakah soal TTMCQ yang dikembangkan menggunakan bahasa yang mudah dipahami oleh mahasiswa atau sebaliknya dan mengetahui apakah waktu yang diberikan dirasa cukup untuk mengerjakan soal yang diberikan. Hasil dari uji skala terbatas ini digunakan sebagai dasar untuk merevisi butir soal untuk uji coba selanjutnya.

Uji coba skala terbatas dilaksanakan pada mahasiswa STAB Negeri Raden Wijaya Wonogiri Jawa Tengah dimana uji skala terbatas untuk mengetahui tanggapan untuk soal yang dikembangkan dapat disimpulkan sebagai berikut: hampir seluruh siswa peserta uji coba skala terbatas belum pernah mengerjakan soal dengan model TTMCQ. Namun dilihat dari soal yang diberikan, sebagaimana besar mahasiswa menyatakan bahwa soal-soal yang diberikan cukup dapat dipahami, hanya pada beberapa soal terdapat data yang kurang lengkap dan terdapat beberapa soal yang tidak memiliki jawaban pada option yang diberikan, sehingga diperlu-

kan beberapa perbaikan pada soal-soal tersebut. dilihat dari tingkat kesukaran soal, terdapat beberapa soal yang tergolong sukar sehingga dalam pengerjaannya membutuhkan waktu yang lebih lama dibandingkan dengan soal yang lain. Sedangkan dari segi waktu yang diberikan yaitu 30 menit dirasa cukup untuk mengerjakan 15 soal TTMCQ yang diberikan.

Uji Coba Skala Menengah

Tahap uji coba selanjutnya adalah uji coba skala menengah. Uji coba ini dilakukan dengan mengujikan soal two-tier ke satu kelas pada masing-masing sekolah. Dimana tujuan dari dilaksanakannya uji coba ini adalah untuk mengetahui karakteristik soal TTMCQ yang dikembangkan.

Suatu soal juga dapat dikatakan sebagai soal yang baik apabila memiliki sifat keajegan atau konstan, atau disebut juga reliabel. Reliabilitas atau keajegan suatu soal ditunjukkan dalam bentuk derajat keajegan yang bernilai antara 0-1. Dimana suatu soal dikatakan reliabel, apabila derajat keajegan minimal adalah 0,700. Hasil koefisien reliabilitas untuk soal utama (*first tier*) adalah 0,754 dan soal alasan (*second tier*) adalah 0,826. Hal ini berarti bahwa soal two-tier yang dikembangkan memiliki tingkat kesukaran yang baik (sesuai dengan tujuan yang diinginkan).

Tingkat kesukaran suatu butir soal berfungsi sesuai dengan tujuan dari tes. Dimana untuk soal TTMCQ yang dikembangkan ini berfungsi sebagai tes formatif, sehingga digunakan butir soal dengan proporsi tingkat kesukaran mudah, cukup dan sukar. Hasil pengujian tingkat kesukaran soal dari uji skala menengah yang telah dilakukan untuk soal utama (*first tier*) yaitu untuk kategori Mudah 40%, Cukup 53,33 % dan Sukar 6,67, sedangkan untuk soal alasan (*second tier*) untuk kategori Mudah 60%, Cukup 33,3% dan Sukar 6,67% sehingga soal TTMCQ dikategorikan sebagai soal yang cukup.

Salah satu syarat kriteria soal yang baik, selain dilihat berdasarkan reliabilitas dan tingkat kesukarannya, dilihat juga berdasarkan daya pembeda soal. Hasil perhitungan daya pembeda soal hasil uji coba skala menengah diperoleh bahwa daya pembeda soal, baik soal utama (*first tier*) maupun soal alasan (*second tier*) memiliki daya beda yang melebihi 0,2 berdasarkan Crocker dan Algina (dalam Yamtinah, 2009), dikatakan bahwa apabila suatu butir soal memiliki daya beda lebih dari 0,2 berarti bahwa soal dapat digunakan. Selain itu, daya pembeda soal seluruhnya

bernilai positif, hal ini menandakan bahwa soal yang dikembangkan telah membedakan siswa kelompok atas (siswa skor tinggi) dengan siswa kelompok bawah (siswa dengan skor rendah). Hasil yang diperoleh dari daya pembeda soal untuk soal utama kategori soal Jelek 13,33%, Cukup 13,33%, Baik 66,67% dan Baik Sekali 6,67% sedangkan untuk soal alasan kategori soal Cukup 26,67%, Baik 46,66% dan baik Sekali 26,67% yang kesemuanya disimpulkan dari daya pembeda soal ini dikategorikan soal baik.

Uji Coba Skala Luas

Tahap uji coba yang terakhir adalah uji coba skala luas. Uji coba skala luas dilakukan dengan mengujicobakan soal kedua sekolah. Tujuan dari dilakukan uji skala luas ini adalah untuk memperoleh tanggapan terkait dengan soal TTMCQ yang dikembangkan.

Soal TTMCQ pada tahap ini diujikan kepada siswa kelas 7 di SMP Negeri 1 Keling dan SMP Negeri 2 Keling. Uji coba dilaksanakan sebanyak tiga kali. Dimana berdasarkan hasil perhitungan reliabilitas 15 butir soal yang dikembangkan dapat dikatakan layak digunakan karena memberikan harga 0,754.

Pada tahap ini juga diberikan angket tanggapan kepada siswa terkait dengan soal yang dikembangkan, khususnya soal TTMCQ untuk materi Pancadharma. Adapun hasil dari tanggapan guru berkaitan dengan soal adalah bahwa soal TTMCQ merupakan soal yang baik. Hanya saja penggunaannya yang masih sangat jarang disekolah-sekolah, sehingga diperlukan sosialisasi berkaitan dengan soal dengan model TTMCQ. Selain itu, soal yang dikembangkan tergolong mudah dan dilihat dari segi waktu 30 menit untuk 15 butir soal yang dirasa cukup, namun karena siswa atau siswa belum terbiasa dengan model TTMCQ ini, maka untuk siswa yang tergolong ke dalam siswa kelompok bawah membutuhkan waktu yang lebih lama, sehingga diperlukan pertimbangan kembali berkaitan waktu yang diberikan dan tingkat kesukaran soal.

Respon Guru Terhadap Instrumen Two-Tier Multiple Choice Question

Respon guru terhadap instrument penilaian hasil pengembangan diperoleh dari hasil angket dan wawancara. Hasilnya antara lain guru belum pernah menggunakan soal dalam bentuk *two-tier multiple choice question* (pilihan ganda bertingkat), soal dalam bentuk pilihan ganda bertingkat lebih rumit dibandingkan dengan pilihan ganda biasa, bentuk

soal suah bagus dan tertata rapi, bentuk soal ini lebih menantang kemampuan berpikir siswa, soal ini cocok untuk melihat kemampuan berpikir dari siswa dan sejauh mana siswa dalam menyelesaikan, soal ini cocok digunakan untuk melihat efektivitas pembelajaran yang dilakukan guru serta pemahaman konsep pada siswa terhadap materi yang diajarkan.

Hasil angket menunjukkan bahwa guru bersedia menggunakan kembali soal pilihan ganda bertingkat untuk mengukur pemahaman siswa pada materi Pancadharmas. Bentuk soal ini dinilai lebih cocok sebagai soal diagnosis keterampilan berpikir. Soal jenis ini menantang untuk siswa untuk meningkatkan kemampuan berpikir yang tidak hanya hafalan tetapi juga pemahaman akan konsep, dapat digunakan sebagai soal ulangan harian karena dapat focus untuk materi apa saja dan soal termasuk jenis soal sederhana tetapi siswa harus berpikir untuk memahami konsep dalam materi dan menghindari siswa untuk menebak jawaban seperti jika mengerjakan soal pilihan ganda biasa karena adanya tingkatan soal yaitu soal utama dan soal alasan yang saling terkait.

Respon guru terhadap instrument penilaian hasil pengembangan sesuai dengan penelitian Tuysuz (2009) yang sejalan dengan hasil penelitian membuktikan bahwa penggunaan TTMCQ dapat membantu guru untuk mengajar lebih baik dan membantu siswa untuk belajar lebih baik. Bentuk soal TTMCQ dapat digunakan oleh guru untuk meningkatkan pengetahuan siswa dan menjaga pemahaman alternative siswa serta membantu meningkatkan proses belajar mengajar apabila dibandingkan dengan tes pilihan ganda biasa. Hasil penelitian Nofiana (2013) juga mengungkapkan bahwa penerapan instrument penilaian evaluasi *two-tier multiple choice question* untuk mengukur keterampilan berpikir tingkat tinggi yang dirancang sedemikian rupa dapat memberikan pengaruh yang positif atau signifikan terhadap hasil belajar siswa.

Kesimpulan

Instrumen penilaian Two Tier Multiple Choice Question (TTMCQ) pada materi Pancadharmas yang dikembangkan layak dan memenuhi kriteria sebagai soal yang baik, dimana dari penelitian yang telah dilakukan diperoleh validasi isi 92,22%; reliabilitas untuk *first tier* 0,754 dan *two tier* 0,826; daya pembeda soal dikategorikan soal dalam kategori soal yang baik untuk soal utama 66,67% dalam kategori baik dan soal alasan

46,66% dalam kategori baik; serta tingkat kesukaran 40% Mudah dan 53,33% Sedang, lalu untuk soal alasan 60% dalam kategori mudah dan 33,33% dalam kategori sedang. Tanggapan guru dan siswa terhadap instrument TTMCQ yang dikembangkan yang diperoleh dari hasil angket dan wawancara, diperoleh hasil antar lain guru belum pernah menggunakan soal dalam bentuk two-tier multiple choice question, soal dalam bentuk pilihan ganda bertingkat lebih rumit dibandingkan dengan pilihan ganda biasa, bentuk soal sudah bagus dan tertata rapi, bentuk soal lebih menantang kemampuan anak dalam berpikir dan menyelesaikan soal, soal ini cocok digunakan untuk melihat efektivitas pembelajaran yang dilakukan guru serta pemahaman konsep pada siswa terhadap materi yang diajarkan. Hasil angket menunjukkan bahwa guru bersedia menggunakan kembali soal pilihan ganda bertingkat dengan alasan bentuk soal ini menantang siswa dalam berpikir, dapat digunakan sebagai soal ulangan harian untuk semua materi.

Daftar Pustaka

- Adodo, S.O. 2013. *Effect of Two-Tier Multiple Choice Diagnostic Assesment Items on Student's Learning Outcome in Basic Science Technology. Ondo State: Academic Journal of Interdisciplinary Studies by MCSER-CEMAS-Sapienza University of Rome. E-ISSN 2281-4612. ISSN 2281-3993, Vol No. 2.*
- Alison Cullinane. 2011. *Two-Tier Multiple Choice Question (MCQs)– How effective are they: A pre-service teachers' perspective.* IOSTE-NW Europe.
- Ardiansah, Masykuri, and Raharjo. 2018. *Student Certainty Answering Misconception Question: Study of Three-Tier Multiple Choice Diagnostic Test in Acid-Base and Solubility Equilibrium.* Journal of Physics: Conference Series Volume 1006 Issue 1 article id. 012018 (2018)
- Arikunto, Suharsimi. 2006. *Dasar-Dasar Evaluasi Pendidikan.* Jakarta: PT. Bumi Aksara
- Arslan, Harikan Ozge, Christine. 2012. *A Three-Tier Diagnostic Test to Assess Pre-Service Teachers' Misconceptions about Global Warming, Greenbaouse Effect, Ozone Layer Depletion and Acid Rain.* Interntional Journal of Science Education Volume 34 n11

p1667-1686.

- Beyza Karadeniz Bayrak. 2013. *Using Two-Tier to identify Primary Students' Conceptual Understanding and Alternative Conceptions in Acid Base*. Mevlana International Journal of education (MIJE) Volume 3(2) pp. 19-26
- Borg and Gall. 1983. *Education Research an Introduction*. New York: Logman Inc
- Caleon, Imelda and Subramaniam. 2010. *Do Students know What They Know and What They Son't Know? Using a Four Tier Diagnostic Test to Assess the Nature of Student Alternative Conceptions*. Journal Research in Science Education Volume 40 (3) pp 313-337.
- Chandrasegaran, A.L., Treagust, D. F Mocerino, M. 2007. *The Development of a Two-Tier Multiple-Choice Diagnostic Instrument for Evaluating Secondary School Student's Ability to Describe and Explain Chemical Reactions Using Multiple Level of Representation*. The Royal Society of Chemistry: Chemistry Education Research and Practice, Vol 8. (3), 293 - 307
- Howell, Dusti, Tseng, Daphine Ching Yu, Colorado-Resa, Jozenia. 2017. *Fast Assessment with Digital Tools Using Multiple-Choice Questions*. Journal Collage Teaching Volume 65 (3) pp 145-147.
- Kuo, Bor-Chen, Magdalena Mo Ching. 2016. *Cognitive Diagnostic Model for Tests with Multiple Choice and Contructed-Response Items*. Kournal Educational Psychology Volume 36 (6) pp 1115-1133.
- Lin, Jing-Wen. 2016. *Development and Evaluation of the Diagnostic Power for a Computer-Based Two Tier Assesment*. Journal of Scince Education and Technology Volume 25 (3) pp 497-511.
- Loh, Andrian Sin Loy, Subramaniam, Tn, Kim Chwee Daniel. 2014. *Exploring Students' Understanding of Electrochemical Cells Using an Enhanced Two-Tier Diagnostic Instrument*. Research in Science and Technology Education, V32 (3) p229-250 2014.
- Maulita, Sukarman dan Marzuki. 2019. The Content Validity: Two-Tier Multiple Choice Instrument to Measure Higher-Order Thingking Skills. YSSTEE 2018 IOP Publishing, IOP Conf. *Series Journal of Physics Conf. Series* 1155 (2019) 012042 doi 10.1088/1742-6596/1155/1/012042.
- Musfiqon. 2012. *Pengembangan Media dan Sumber Media Pembelajaran*. Jakarta: PT. Prestasi Pustakaraya
- Nofiana. Mufida 2014. Pengembangan Instrumen Evaluasi *Two-Tier*

- Multiple Choice Question* untuk Mengukur Keterampilan Berpikir Tingkat Tinggi. *Jurnal FKIP UNS dalam Prosiding Seminar Nasional Pendidikan Sains IV 2014* ISSN: 2407-4659
- Siswaningsih, Firman and Zackiyah. 2017. Development of Two-Tier Diagnostic Test Pictorial-Based for Identifying High School Student Misconceptions on the Mole Concept. *Journal of Physics Conference Series Volume 812*, (1) article id. 012117.
- Smith, Mark. 2017. Cognitive Validity: On Multiple-Choice Items Tap Historical Thinking Processes. *American Educational Research Journal Volume 54* (6) pp 1256-1287.
- Sugiyono. 2012. *Metode Penelitian Pendidikan: Pendekatan Kuantitatif, Kualitatif dan R&D*. Bandung: Alfabeta
- Tsui, Chi-Yan, Treagust, David. 2010. Evaluating Secondary Students' Scientific Reasoning in Genetics Using a Two-Tier Diagnostic Instrument. *Internasional Journal of Science Education Volume 32* (8) pp 1073-1098.
- Tuysuz, Cengiz. 2009. Development of Two Tier Diagnostic Instrument and Assess Students Understanding in Chemistry. *Scientific Research and Essay Volume 4* (6) pp 626-631
- Widiyatmoko and Shimizu. 2018. The Development of Two-Tier Multiple Choice Test to Assess Student' Conceptual Understanding About Light and Optical Instrument. *JPII 7* (4) (2018) 491-501 DOI: 10.15294/jpii.v7i4.16591.
- Winarti, Cari, Suparmi, Sunarno, Widha, Istiyono, Edi. 2017. Development of Two Tier Test to Assess Conceptual Understanding in Heat and Temperature. *Journal of Physics: Conference Series. Volume 795* (1) Article id. 012052.
- Yang, Tzu-Chi, Fu, Hseng-Tz, Gwo-Jen, Yang, Stephen. 2017. Development of an Interactive Mathematics Learning System Based on a Two-Tier Test Diagnostic and Guiding Strategy. *Australasian Journal of Education Technology Volume 33* (1) pp 62-80.