

BAB I

PENDAHULUAN

A. Latar Belakang

Penilaian merupakan bagian penting dan aspek yang tidak terpisahkan dalam penyelenggaraan pendidikan. Sementara itu tes digunakan dalam penilaian dalam bentuk instrumen dan informasi yang dikumpulkan berupa karakteristik dari suatu objek (Widoyoko, 2015). Tes yang digunakan untuk mengetahui sejauh mana tingkat pemahaman siswa setelah mengikuti proses belajar mengajar adalah *Assesment of Learning* selanjutnya disebut AoL (Azwar, 2012).

Tes prestasi dikelompokkan menjadi dua jenis, yaitu tes objektif dan tes subjektif. Tes objektif contohnya adalah pilihan ganda, benar salah dan jawaban singkat sementara tes subjektif contohnya uraian (Affiong, Nkwol dan Nsungo, 2014). Ujian Nasional selanjutnya disebut UN dilakukan oleh pemerintah dan berstandar Nasional yang disiapkan seluruh butirnya oleh kementerian pendidikan dan kebudayaan selanjutnya disebut Kemendikbud (Cahyo, Fariz, dan Lestari, 2020) merupakan tes prestasi hasil belajar (Hidayah, 2013) dan UN termasuk AoL dalam tes objektif, sebab jenis soalnya pilihan ganda dan isian singkat (Istiyono, Mardapi, dan Suparno, 2014).

UN dalam mengevaluasi dan menilai prestasi hasil belajar peserta didik dari setiap proses dalam pendidikan dasar dan menengah di Indonesia (Idrus, 2019). Namun setiap penyelenggaraan UN terdapat kesamaan dalam hal desain tesnya yaitu seluruh penempuh tes (tanpa melihat kemampuannya) dalam jenjang pendidikan sama memperoleh butir-butir soal yang relatif sama (setara tingkat kesulitannya) dan pembuatannya dalam satu kisi-kisi (Hamiseno, Bagus, Fahmi, Zulaiha, Susetyo & Rahman, 2020). Selain itu juga, paket soal yang ditentukan untuk kebutuhan alat ukur sampai saat ini belum

memperhatikan kemampuan penempuh tes secara unik (Rukli, 2013). Pengabaian terhadap variasi kemampuan individu tersebut merupakan kelemahan dari pendekatan teori klasik oleh karena semua peserta tes akan mendapatkan seperangkat item yang sama atau yang setara (Embretson dan Reise, 2000).

Berdasarkan fakta yang ada bahwa setiap penempuh tes memiliki variasi kemampuan yang signifikan berbeda. Sejalan dengan konsep tidak ada dua orang di dunia ini yang benar-benar sama dalam segala hal, sekalipun mereka kembar (Nurdin, 2005). Dengan mengabaikan variasi kemampuan individu dalam model penilaian tersebut terdapat kelemahan, hal ini menurut Binet dalam Linacre (2000) tidak ada keadilan pada penyelenggaraan tes yang memberikan butir soal yang sama kepada semua penempuh tes. Sehingga butir yang terlalu mudah tidak mempengaruhi siswa untuk dapat berusaha dalam memecahkan butir tersebut sedangkan butir yang terlalu sulit dikerjakan maka siswa akan frustrasi sebab untuk dapat memecahkan butir itu diluar dari kemampuannya kemudian tidak lagi termotivasi dan bersemangat untuk menyelesaikannya (Halik, Mania, dan Nur, 2019).

Sebagai gambaran ketika ujian mengabaikan variasi kemampuan penempuh tes yaitu jika paket ujian dengan tingkat kategori butir mudah dan diberikan kepada sekelompok penempuh tes dengan tingkat kemampuan kognitifnya tinggi maka paket tersebut memberikan informasi yang kurang berarti dan penempuh tes tersebut menjadi tidak tertantang karena dapat dipastikan hasil diperoleh yang maksimal, karena semua penempuh tes akan menjawab benar pada paket tersebut. Sebaliknya, jika paket dengan kategori butir sulit diberikan kepada sekelompok penempuh tes dengan kemampuan kognitifnya rendah maka kemungkinan semua penempuh tes memperoleh skor nol atau minimum dan penempuh tersebut menjadi frustrasi, karena tidak dapat menjawab seluruh butir soalnya.

Butir soal pada AoL dikatakan baik apabila butir yang diberikan tersebut tidak terlalu mudah dan juga tidak terlalu sulit (Halik et al,2019). Apabila Butir soal dikaitkan dengan tingkat kesukaran maka butir soal tersebut sangat dipengaruhi oleh kemampuan dari kompetensi terhadap literasi (*literacy reading*) dan numerasi (*literacy numeric*) dari penempuh tes yang mengerjakan (Howard, Woodcock,Ehrich, dan Bokosmaty, 2016). Selain itu juga tingkat kesukaran butir juga dipengaruhi oleh konstruk validitas dan reliabilitas butir serta waktu respon yang dibatasi dalam menjawab butir tersebut (Sukardi,2008).

Berkaitan dengan waktu respon yang dibatasi tersebut juga menjadi permasalahan yang dihadapi berikutnya. Hal ini berdasarkan POS UN (2019), alokasi waktu untuk seluruh mata ujian yaitu 120 menit dengan model tesnya pilihan ganda sebanyak 40 butir soal. Sehingga selama periode waktu 120 menit tersebut penempuh tes harus selesai mengerjakan tes, apabila ada butir yang belum direspon maka secara otomatis sistem akan mencatat sebagai respon jawaban salah dan memperoleh skor nol. Menurut Nitko (1996) memperkiraan waktu respon butir untuk model soal pilihan ganda yaitu antara 40 hingga 60 detik, sementara model tes isian singkat setiap butir memiliki perkiraan waktu respon butir yaitu antara 15 sampai dengan 20 menit. Sehingga dengan pendekatan Nitko (1996) untuk menyelesaikan 40 butir membutuhkan periode waktu antara 27 menit hingga 40 menit, terdapat kelebihan 80 menit.

Wainer (2000) berasumsi bahwa pada tes yang dibatasi oleh waktu, penempuh tes tersebut menjadi tidak memiliki performance dengan baik, namun Bodman dan Robinson (2004) menyatakan bahwa motivasi terhadap ujian berkorelasi negatif terhadap waktu pengujiannya, atau semakin lama tes yang diberikan semakin tidak termotivasi. Sementara itu ujian dengan resiko tinggi atau *high stake* dibatasinya waktu penempuh dalam merespon butir menjadi penting sebab diasumsikan penempuh tes memiliki

kecepatan yang konstan selama pengujian (Linden, 2007). Sejalan dengan hal tersebut maka UN memang dapat dibatasi waktu sebab UN dianggap sebagai *high-stakes testing* karena aspek kejujurannya rendah (Saukah dan Cahyono, 2015).

Berkaitan dengan dibatasi waktu dalam ujian, Hall (1987) membagi konsep waktu menjadi dua, yakni waktu monokronik dan waktu polikronik. Budaya pengikut waktu monokronik mengaitkan akan waktu yang berjalan terus dari masa lalu ke masa yang akan datang sehingga mereka sangat menekankan sekali terhadap harga akan waktu. Sebaliknya, pengikut waktu polikronik mengaitkan waktu sebagai putaran yang kembali dan kembali lagi secara berulang. Hal inilah yang mengakibatkan mereka akan memperlakukan waktu lebih santai.

Indonesia cenderung menganut konsep waktu polikronik. Hal ini dikarenakan seringkali masyarakat menunda-nunda pekerjaan sampai akhirnya terlambat. Fenomena ini menjadi suatu kebiasaan yang membudaya pada masyarakat Indonesia (Mulyana, 2004). Namun masyarakat pengikut polikronik dengan diberikan deadline waktu maka mereka akan berkinerja lebih baik dibandingkan dengan penganut monokronik disaat banyak tugas yang dikerjakan (Guise, Knott, dan Benuskova, 2014). Terkait dengan etnographi di Indonesia yang masuk kategori polikronik tersebut di atas agar kinerjanya menjadi lebih baik dari monokronik, jika dikaitkan dengan AoL sebagai tes prestasi belajar maka perlu adanya pembatasan waktu pengerjaan tes.

Berdasarkan penjelasan di atas, cara mengatasi kelemahan dalam pengadministrasian AoL sebagai tes prestasi hasil belajar berskala Nasional atau UN yang mengabaikan variasi kemampuan penempuh tes, harus dilakukan penelitian model penilaian yang ideal, sehingga karakteristik butir yang dikerjakan oleh penempuh tes akan disesuaikan dengan kemampuan penempuh yang bervariasi. Sejalan dengan Davey (2011) yang menjelaskan bahwa penentuan target agar butir yang diberikan ke penempuh

tes disesuaikan dengan kemampuannya maka akan menghasilkan sedikit kesalahan dalam tes. Cara tersebut menghasilkan paket soal lebih spesifik dan *on target* kearah kemampuan penempuh tes. Untuk dapat memperoleh paket soal yang spesifik kearah kemampuan penempuh tes tersebut dapat dilakukan dengan model tes adaptif.

Dewasa ini tes adaptif sudah dibantu dalam penyajiannya menggunakan komputer, atau selanjutnya dapat disebut *Computerized Adaptive Testing* selanjutnya disingkat CAT. Dalam CAT, kemampuan penempuh tes diperkirakan secara iterasi atau hitungan berulang setelah setiap item tes dijawab, dan item berikutnya dipilih berdasarkan estimasi kemampuan saat ini (lord, 1977). Tujuan dari CAT menurut Riza, Nurjanah, dan Wihardi (2019) yaitu proses merakit tes dengan bangkitan butir soalnya disesuaikan dengan kemampuan penempuh tesnya, dengan kemampuan penempuh tesnya dapat diestimasi setelah butir di respon oleh penempuhnya. Sehingga menurut Lord (1977) apabila penempuh tes memiliki tingkat kemampuan tinggi maka butir yang diberikan ialah butir dengan tingkat kesukaran tinggi sedangkan butir yang mudah tidak diberikan. sehingga, penempuh tes tidak lagi diberikan seluruh butir. Hal ini lah memungkinkan waktu pengetesan dengan moda tes adaptif akan lebih singkat dibandingkan dengan moda konvensional dimana desain yang sama untuk seluruh penempuh tesnya (Bagus, 2012). Selain itu, Bagus (2012) mengutarakan bahwa model CAT menghasilkan *Standard error of Measurement (SEM)* lebih kecil karena butir yang dibangkitkan disesuaikan dengan estimasi kemampuan penempuhnya.

Berdasarkan penjelasan kelemahan UN di atas, model CAT dapat mengestimasi performa seseorang atau sekelompok orang dalam sebuah butir. Menurut Retnawati (2014) analisis terhadap estimasi performa seseorang atau sekelompok orang dalam sebuah butir tanpa tergantung lagi oleh kelompok tes dan paket tes maka digunakan alternatif pendekatan modern atau *item respon theory (IRT)* . Selain itu pendekatan IRT

meletakkan sebuah dimensi pengetahuan atau sifat dimana butir tes bergantung pada respon benar dari individu penempuh tes (Wainer, dan Mislevy, 1990). Posisi setiap butir pada dimensi tersebut diketahui sebagai kesukaran butir (b) dan posisi setiap penempuh tes menjawab benar diketahui sebagai *proficiency* atau *ability* (θ). Menurut Embretson dan Reise (2000), kemampuan penempuh tes dan kesukaran butir berada pada skala yang sama. Karenanya penempuh tes besar kemungkinan mendapatkan seperangkat butir yang berbeda dengan penempuh tes yang lain (Wainer, 1990).

Keunggulan lain terhadap pelaksanaan CAT yaitu mudah dalam pengadministrasian tesnya dan menurut Haar dan Hansen (2006) siswa lebih menikmati dan termotivasi dengan menggunakan moda tes terkomputerisasi atau CAT daripada Ujian berbasis kertas (UKP), dan siswa lebih fokus terhadap ujian, lebih menikmati, serta lebih merasa mampu mengerjakan ujian dengan moda tes Ujian Berbasis Komputer selanjutnya disebut UBK (Morgan dan O'Reilly, 2001) dan (Solomon, 1949). Menurut pramono (2020) CAT bersifat adaptif, sehingga butir yang diberikan tidak ada butir yang sangat sulit atau sangat mudah untuk penempuhnya. Model ini memungkinkan informasi yang diperoleh *on target*, yaitu tes dengan bangkitan satu butir tingkat kesulitan sesuai kemampuan penempuhnya, begitu juga butir berikutnya, dan model CAT akan berhenti ketika estimasi kemampuan penemuhnya telah konvergen dengan hasil kesalahan tertentu. Hal ini menghasilkan waktu yang dibutuhkan untuk ujian menjadi singkat.

Menjadi menarik untuk diteliti yaitu ketika AoL terstandar Nasional atau UN dalam administrasi tesnya dimodelkan dengan tes adaptif berbasis komputer atau CAT dan dengan membatasi waktu respon butir. Secara operasional umum dari CAT menurut Choe, Kern & Chang (2017) mengatakan efisiensi pengukuran estimasi kemampuan dengan model CAT tidak hanya dinilai dari banyaknya butir yang diberikan tetapi juga waktu yang dibutuhkan untuk menyelesaikan tes juga diperhatikan. Hal ini dalam

penelitian memanfaatkan data respon dan data waktu respon butir dari UNBK. Dengan data-data tersebut akan dilakukan pengujian ulang untuk mengestimasi kemampuan penempuh tes dan merakit tes dengan pemodelan administrasi CAT, dan hasilnya akan dibandingkan dengan administrasi UNBK.

Untuk dapat memperoleh data penelitian ini, diperoleh data dari penyelenggaraan UN Berbasis Komputer selanjutnya disebut UNBK berupa log pencatatan aktifitas pengerjaan siswa yang informasinya terekam secara cepat oleh sistem yang sudah terkomputerisasi (Mardiana dan Handyani, 2017). Oleh karena itu aktifitas siswa dari awal hingga akhir tes tercatat dan terekam oleh sistem seperti respon siswa, waktu respon butir, kunjungan/singgah butir soal (*visit*) dari butir satu ke butir yang lain, dan beberapa aktifitas lainnya. Hal ini sejalan dengan yang disampaikan van der Linden (2006) yaitu dengan kedatangan pengujian berbasis komputer menjadikan pencatatan waktu respons sebagai bagian rutin dari moda tes.

Hasil perekaman serta pencatatan aktifitas siswa tersebut, terdapat dua hal informasi yang penting untuk di jadikan data penelitian yaitu respon butir siswa dan waktu respon butir siswa. Informasi pertama yaitu respon butir siswa merupakan data penting dalam penelitian sebab dipakai untuk mengestimasi tingkat kesukaran butir dan estimasi kemampuan penempuh tesnya dengan pendekatan teori modern (*Item Response Theory*) selanjutnya disebut IRT. Informasi penting kedua yaitu waktu respon butir setiap penempuh dalam merespon butir.

Penelitian terbaru tentang pembatasan waktu dalam tes atau ujian dikaitkan dengan estimasi parameter (Bolt, Cohen, & Wolack, 2002; Goegebeur, DeBoeck, Wollack, & Cohen, 2008), pengaruh terhadap validitas (Lu & Sireci, 2007), dan kemungkinan untuk mencegahnya (van der Linden & Xiong, 2013). Oshima (1994) mengatakan bahwa secara fungsional terdapat hubungan dengan karakteristik peserta dan adanya pengaruh

lainnya sehingga rentang waktu yang diberikan berapapun mungkin tidak cukup untuk peserta yang sangat lambat, dan dalam framework IRT disebut sebagai kecepatan. Untuk mengatasi kelalaian terhadap batas waktu tersebut Kubinger (2016) dan van der Linden dan Xiong (2013) menyarankan administrasi tesnya menggunakan model CAT.

B. Pembatasan Penelitian

Mengingat cukup luasnya permasalahan dalam penelitian ini dan keterbatasan yang penulis miliki, maka masalah yang akan dibahas dalam penelitian ini dibatasi sebagai berikut:

1. Administrasi tes menggunakan model CAT dengan pendekatan Maksimum Likelihood dalam mengestimasi kemampuan penempuh tes pada teori tes modern IRT.
2. *Content balancing* dan *Exposure Control* tidak diikuti dalam penelitian ini.
3. Fokus pengembangan yaitu data respon siswa UNBK yang dimodelkan dengan model CAT dan penerapan konstrain waktu respon butir sebagai batas penempuh tes dalam merespon setiap butir soalnya.
4. Data penelitiannya menggunakan data UNBK tahun 2019 pada propinsi DKI Jakarta jenjang SMP, dan mata pelajaran Matematika
5. Waktu respon butir sebagai batasan waktu penempuh dalam merespon butir hanya dipengaruhi oleh aspek butir penyebab waktu respon lama, misalnya tingkat kesukaran butir dan indeks daya beda penempuh tes. Sedangkan aspek penempuh tes lain penyebab lamanya waktu respon butir misalnya memiliki kelebihan kompetensi literasi, numerasi dan pembiasaan dalam mengerjakan butir.

C. Masalah Penelitian

Sejak tahun 2015 hingga 2020, setiap tahun peserta didik pada jenjang akhir pada pendidikan dasar dan menengah di seluruh Indonesia menghadapi UN. Berdasarkan sifatnya, UN merupakan penilaian hasil belajar secara sumatif atau istilah lain sering disebut sebagai *assessment of learning* (AoL) terstandar Nasional untuk ujian jenjang akhirnya. Desain paket soal UN terdiri atas 100% konstruksi butir disiapkan oleh Kemdikbud.

Pada Penelitian ini mencoba untuk mencari solusi yang tepat pada metode AoL terstandar sebagai tes prestasi hasil belajar dan melakukan analisis mendalam kaitannya dengan waktu respon butir soal serta faktor yang mempengaruhinya. Penelitian semacam ini sudah pernah dilakukan oleh Bagus (2012) namun belum sampai mendalam khususnya peranan waktu respon butir dari pemodelan CAT dibandingkan dengan *Pencil Test* (PPT). Menurut Bagus (2012), AoL terstandar dengan moda pelaksanaan CAT lebih efisien sebab penempuh tes tidak perlu mengerjakan seluruh butir dalam satu paketnya, melainkan CAT akan membangkitkan butir yang sesuai dengan estimasi kemampuan penempuh tesnya, sehingga jumlah butir yang dikerjakan lebih sedikit. Dengan jumlah butir yang lebih sedikit tersebut dikatakan lebih efisien. Untuk penelitian efektifitas yang sudah dilakukan oleh Bagus (2012) dengan membandingkan besarnya Standar Error of Measurement (SEM) antara PPT dan CAT. Dari hasilnya moda CAT memiliki SEM yang lebih kecil dibandingkan dengan moda PPT.

Dalam penelitian disertasi ini nantinya peneliti akan mengkonstrain waktu respon butir soal yang diterapkan dalam pemodelan CAT sehingga waktu respon butir akan dicatat dan diukur serta diuji efisiensinya dan efektifitasnya. Dengan adanya kebutuhan pembatasan periode waktu pengerjaan ini menjadi sangat penting untuk diteliti dan dianalisis lebih mendalam, sebab dengan tidak dibatasinya waktu respon butir soal

tersebut membuat penempuh tes tidak serius dan tidak termotivasi. Oleh karena itu harus ada konstrain waktu respon butir agar penempuh tes menjadi serius dan motivasi tinggi (Bodmen & Robinson, 2004). Sehingga semakin ketat waktu response butir dibatasi maka motivasi dalam mengerjakan ujian semakin tinggi dan penempuh tes akan semakin serius. Desertasi ini akan meneliti faktor yang mempengaruhi waktu respon butir misalnya tingkat kesukaran, daya beda butir, jumlah kunjungan atau jumlah *visit* ke butir soal.

Pada penelitian desertasi ini nantinya peneliti mencoba untuk mengukur dan menganalisis butir-butir hasil bangkitan model CAT yang dikonstrainsi oleh waktu respon butir. Dalam administrasi dengan model CAT setiap kali penempuh tes merespon butir jawaban dan berpindah butir maka sistem model CAT akan mencatat beberapa hal diantaranya tingkat kesukaran butir soal hasil bangkitan, hasil estimasi kemampuan penempuh tes dan waktu respon butir soal. Butir soal hasil bangkitan akan menyesuaikan dengan estimasi kemampuan penempuh tesnya (Lord, 1977), sehingga nanti akan peneliti ukur selisih antara estimasi kemampuan penempuh tesnya dengan butir soal hasil bangkitannya tersebut, dan hasil selisih tersebut akan diuji kepresisian.

Di dalam penelitian ini dilakukan analisis untuk membandingkan estimasi kemampuan penempuh tes apabila model CAT dikonstrainsi waktu respon butirnya dan tanpa dikonstrainsi waktu respon butirnya. Selanjutnya hasil estimasi kemampuan penempuh tes tersebut juga akan dibanding dengan data *true score* atau kemampuan acuan penempuh tes dan akan diuji keakuratan dari kedua kondisi tersebut terhadap *true score*-nya. Dengan uji *Mean Square Error* (MSE) akan dipilih MSE terkecil yang menentukan kondisi model CAT mana (dibatasi waktu respon butir atau tanpa batas waktu respon) yang lebih akurat.

Diharapkan dengan penelitian yang dilakukan ini dapat diperoleh sebuah model penilaian yang ideal, lebih tepat, dan akurat untuk digunakan dalam melakukan penilaian hasil belajar siswa dengan menerapkan teori pengukuran modern di dalam evaluasi Pendidikan.

C. Rumusan Masalah

Berdasarkan pada latar belakang dan masalah masalah sebagaimana telah diuraikan di atas, maka rumusan masalah dalam penelitian ini adalah disusun sebagai berikut:

1. Apakah terdapat perbedaan efisiensi dan efektifitas terhadap AoL terstandar yaitu antara administrasi model UNBK, model CAT yang mengkonstrain waktu respon butir dan model CAT tanpa mengkonstrain waktu respon butir?
2. Apakah terdapat perbedaan presisi terhadap AoL terstandar dari administrasi model CAT dalam bangkitan butir soal yang menyesuaikan dengan hasil estimasi penempuh tesnya yaitu antara model CAT yang mengkonstrain waktu respon butir, model CAT tanpa mengkonstrain waktu respon butir?
3. Apakah terdapat perbedaan akurasi AoL terstandar dalam mengestimasi kemampuan penempuh tes antara administrasi model UNBK, model CAT yang mengkonstrain waktu respon butir dan model CAT tanpa mengkonstrain waktu respon butir?

E. Tujuan dan Manfaat Penelitian

Berdasarkan latar belakang dan rumusan masalah di atas, penelitian ini bertujuan untuk mendeskripsikan fakta baru terhadap faktor yang mempengaruhi pada penerapan waktu response butir yang membatasi penempuh tes dalam merespon setiap butir soal di

administrasi model CAT. Kaitannya dengan waktu respon butir tersebut juga peneliti ingin menganalisis lebih mendalam model yang lebih efisien dan efektif kaitannya pada AoL terstandar antara administrasi model UNBK, model CAT dengan penerapan konstrain waktu respon butir, dan model CAT tanpa penerapan waktu respon butir soal dengan menguji efisiensi, dan efektifitas dari administrasi model tersebut.

Tujuan berikutnya adalah menganalisis model yang lebih presisi dari hasil hitung selisih yang terkecil dari hasil bangkitan butir soal yang disesuaikan dengan estimasi kemampuan penempuh tes terhadap AoL terstandar antara administrasi model CAT yang mengkonstrain waktu respon butir dan tanpa mengkonstrain waktu respon butir soal. Tujuan terakhir dari penelitian disertasi ini adalah menganalisis keakuratan dari hasil hitung selisih terkecil dengan uji MSE antara *true score* atau kemampuan acuan penempuh tes dengan estimasi kemampuan penempuh model UNBK, model CAT penerapan konstrain waktu respon butir, dan model CAT tanpa penerapan waktu respon butir. Dalam analisis keakuratan hasil estimasi kemampuan penempuh tes dari beberapa model administrasi tes juga akan dibagi kedalam beberapa kelompok level kemampuan penempuh tes (kemampuan penempuh sangat tinggi, tinggi, sedang, rendah dan sangat rendah) dan dikelompokkan kedalam kota/kabupaten di propinsi DKI Jakarta. Produk akhir dari penelitian ini dibuat aplikasi CAT dilengkapi dengan buku panduan dan pengenalan model penilaian CAT.

Manfaat dari penelitian ini akan dibagi menjadi dua yaitu manfaat secara praktis dan manfaat secara teoretis. Manfaat secara praktis dari penelitian ini yaitu diperolehnya alternatif dalam AoL terstandar sebagai solusi atas kelemahan model penilaian selama ini. Alternatif administrasi tesnya yaitu berupa tes adaptif model CAT dengan hasilnya yang on target dan dapat dibandingkan antara sekolah hingga skala Nasional dengan

hasil yang cepat, akurat, presisi, efektif dan efisien namun tetap aman, transparan, objektif, dan akuntabel.

Manfaat praktis berikutnya yaitu penggunaan teknologi komputer saat ini memungkinkan ujian dapat dilakukan dimana saja dan kapan saja sehingga dengan kondisi geografis di Indonesia yang berpulau pulau memungkinkan keberadaan AoL terstandar Nasional menjadi lebih hemat sebab tes bisa terpusat, dengan pengiriman data yang cepat dan hasilnya tetap akurat. Hal ini sejalan dengan target pemerintah melalui Kementerian Komunikasi dan Informasi (Kemenkominfo) tahun 2022 seluruh sekolah akan memperoleh akses internet, dan Kemendikbud memiliki program bantuan Teknologi, Informasi dan Komunikasi (TIK) untuk pembelajaran digital untuk seluruh sekolah yang membutuhkan. Sehingga tes adaptif dengan basis komputer fisibel untuk dilakukan di Indonesia.

Sistem model CAT yang dihasilkan dalam penelitian ini sudah siap dan dapat diterapkan langsung untuk tes hasil belajar tingkat akhir skala Nasional. Sebab Pusat Asesmen Pendidikan (PUSMENDIK) sudah memiliki bank butir dengan jumlah butir soal yang banyak dari seluruh jenjang dengan tingkat validitas dan reliabilitas butir yang baik dan terkalibrasi dari beberapa tahun sebelumnya sehingga model CAT dapat diterapkan secara ideal untuk pemetaan kualitas hasil belajar secara Nasional di Indonesia.

Sementara itu manfaat secara teoretis yaitu membantu para ahli pengukuran sebagai sumber bahan penelitian untuk menemukan hal yang baru khususnya pendekatan IRT dengan pemanfaatan model MLE di administrasi tes CAT pada AoL terstandar sebagai tes prestasi belajar di akhir level setiap jenjang dengan penerapan waktu respon butir yang membatasi penempuh tes dalam merepon jawaban butir soal. Sehingga para akademisi serta praktisi pendidikan dapat menambah wawasan dan

pengetahuan dengan adanya penerapan waktu respon butir tersebut terhadap akurasi dan presisi dari pengukuran estimasi kemampuan penempuh tesnya dengan hasil yang on target. Dengan adanya produk akhir berupa aplikasi CAT dilengkapi buku panduan dan pengenalan CAT sebagai bahan edukasi untuk menambah wawasan baru bagi praktisi pendidikan tentang model penilaian CAT.

F. Kebaharuan Penelitian

Kebaharuan penelitian ini dibagi kedalam tiga kategori :

1. Kebaharuan tipe-1 *invention* / Penemuan:

Kebaharuan pada penelitian ini terletak pada sistem administrasi tes yang adaptif menggunakan model CAT dan waktu yang membatasi penempuh tes dalam meresponse setiap butir. Administrasi pada AoL terstandar tidak hanya adaptif dalam bangkitan butir soal tetapi juga adaptif pada waktu respon butir sesuai dengan tingkat kesukaran dari butir yang dibangkitkan oleh model CAT.

2. Kebaharuan tipe-2 *Improvement*/ Pengembangan:

Kebaharuan dari tipe-2 yaitu dengan adanya modifier waktu respon butir yang terukur dalam membatasi penempuh tes merespon butir soal, administrasi tes pada AoL terstandar menjadi lebih efektif terhadap pemilihan butir soal, efisien waktu tes, presisi terhadap hasil bangkitan yang disesuaikan dengan estimasi kemampuan penempuh tesnya dan tetap akurat dari hasil estimasi kemampuan penempuh tesnya.

3. Kebaharuan tipe-3 *refutation*

Hasil penelitian yang membuktikan bahwa AoL terstandar Nasional berupa *Achievement test*, dapat dimodelkan dengan administrasi model CAT. Model CAT terbukti memperoleh informasi estimasi kemampuan penempuh tes yang

on target dengan desain butir soal yang efektif, periode waktu tes yang efisien, kesesuaian hasil bangkitan butir soal dengan kemampuan penempuh tes yang presisi dan hasil estimasi kemampuan penempuh tesnya tetap akurat.

