Analisis Disparitas Skor Tampak dan Estimasi Skor Murni dengan Pengkategorian Acuan Normatif pada Tes Hasil Belajar Siswa

Eviana Hikamudin[1)], Yahya Hairun[2)]
[1)]Pusat Asesmen dan Pembelajaran, Balitbang Kemendikbud
[2)]Program Studi Pendidikan Matematika, Universitas Khairun
*Corresponding Author: *evianapuspendik@gmail.com*

Abstrak. Penelitian ini bertujuan untuk menganalisis disparitas skor tampak dan skor murni dengan menggunakan pengkategorian berdasarkan acuan normatif pada Tes Hasil Belajar Siswa. Data dalam penelitian ini diperoleh dari hasil Ujian Nasional (UN) matematika jenjang SMP tahun 2016/2017 yang dikeluarkan oleh Pusat Penilaian Pendidikan (Puspendik) Kemendikbud. Analisis data dalam penelitian ini didasarkan pada skor hasil ujian nasional SMP mata pelajaran matematika tahun 2016/2017. Responden yang dianalisis dalam penelitian ini sebanyak 250 orang dan banyaknya butir soal yang dianalisis sebanyak 40 butir. Responden yang terpilih dikelompokkan menjadi 5 grup yang masing-masing berjumlah 50 repsonden dengan jumlah butir yang sama sebanyak 40 butir. Metode yang digunakan adalah metode penelitian kuantitatif dengan pendekatan teori klasik. Pengolahan data yang dilakukan adalah dengan melakukan estimasi terhadap skor murni (T) dengan menggunakan persamaan regeresi. Pengkategorian skor dilakukan berdasarkan acuan normatif dengan memperhitungkan standar eror pengukuran (SEM) untuk meningkatkan kecermatan. Kategori skor yang dihasilkan adalah sebanyak tiga kategori yaitu rendah, sedang, dan tinggi. Hasil analisis data menunjukkan bahwa terdapat disparitas skor tampak (X) dan skor murni (T) setelah dilakukan estimasi terhadap skor murni (T) berdasarkan pengkategorian secara normatif pada masing-masing kelompok. Disparitas skor tampak (X) dan skor murni (T) paling tinggi terdapat pada grup yang memiliki koefisien reliabilitas tesnya paling rendah, sedangkan pada grup yang koefisien reliabilitasnya paling tinggi, disparitas skor tampak (X) dan skor murni (T) paling rendah. Dari hasil penelitian ini dapat disimpulkan bahwa koefisien reliabilitas tes dantingkat disparitas pengkategorian berdasarkan acuan normative menentukan kualitas alat tes.

Kata kunci: e*stimasi skor murni, SEM, disparitas, koefisien reliabilitas*

A. Pendahuluan

Ujian (tes) merupakan bagian yang tidak terpisahkan dari proses pembelajaran. Keberhasilan suatu proses pembelajaran dapat diukur salah satunya melalui ujian. Perancangan ujian perlu dilakukan dengan baik dan terukur agar dapat menghasilkan informasi yang bermanfaat bagi perbaikan mutu pembelajaran. Menurut Naga (1992:1), ujian menyangkut tiga bagian kegiatan penting yaitu konstruksi alat (instrumen) ujian, pelaksanaan ujian oleh peserta ujian (responden), dan penganalisisan butir (item) ujian yang digunakan. Ketiga bagian tersebut merupakan tahapan yang saling berhubungan dan satu sama lain akan menentukan kualitas dari ujian yang dilaksanakan. Instrumen ujian yang berupa seperangkat butir soal harus disusun sesuai dengan kriteria (kisi-kisi) agar dapat mengukur indikator yang telah ditetapkan. Selanjutnya butir soal yang telah

disiapkan harus diujikan kepada responden dengan cara obyektif agar data yang diperoleh dapat menggambarkan kondisi responden yang sebenarnya. Kemudian pada tahap akhir, data respon yang telah diperoleh perlu diolah dan dianalisis dengan baik agar menghasilkan informasi yang akurat untuk dijadikan bahan pengambilan keputusan.

Di bidang pengukurran pendidikan, skor hasil ujian merupakan komponen penting untuk diperhatikan. Skor hasil ujian dapat memberikan informasi mengenai capaian kuantitas maupun kualitas dari ujian yang telah dilakukan. Informasi yang diperoleh melalui skor hasil ujian tidak hanya dapat menjelaskan mengenai karakteristik perserta ujian (responden) tetapi dapat pula menjelaskan tentang karakteristik butir (item) yang digunakan dalam ujian tersebut. Karakteristik responden adalah kemampuan (*ability*) responden dalam merespon butir, sedangkan karakteristik butir dapat berupa tingkat kesukaran (*difficulty*) dan daya pembeda.

Faktor yang selalu terjadi pada setiap pengukuran adalah kekeliruan (*error*). Secara empiris tidak semua hasil pengukuran memiliki tingkat akurasi (ketepatan) yang tinggi sehingga menyebabkan munculnya kekeliruan (*error*) pengukuran. Naga (2013:188) menyebutkan banyak hal yang dapat menyebabkan kekeliruan di dalam pengukuran. Kekeliruan dapat terjadi pada responden, pada pelaksanaan pengukuran, dan pada alat ukur. Pada saat pengukuran dilakukan, kondisi responden tidak selalu mantap dan baik. Gangguan dapat muncul tidak hanya pada diri reponden tetapi dapat muncul dari gangguan fisik atau lingkungan sekitar. Adakalanya alat ukur yang digunakan mengandung keraguan atau ketidakjelasan bagi responden. Kondisi-kondisi seperti itulah yang berpotensi menghasilkan kekeliruan dalam pengukuran.

Bagian yang tidak terpisahkan dari setiap kegiatan pengukuran adalah komponen yang berupa skor. Menurut Azwar (2015:68), performansi individu yaitu respons subyek terhadap item-item dalam skala pengukuran atau tes psikologi, dinyatakan dalam bentuk angka yang disebut dengan skor (*score*). Salah satu sumber yang dapat menghasilkan skor adalah ujian (tes). Skor yang dihasilkan melalui ujian dapat berupa angka (kuantitatif) atau berupa kategori (kualitatif). Skor yang berupa angka biasanya merupakan skor mentah (*raw score*) dan belum dapat menjelaskan performansi individu secara lengkap. Sumintono dan Widhiarso (2015:14) menjelaskan bahwa skor mentah pada dasarrnya bukan hasil pengukuran tetapi merupakan jumlah jawaban benar dari soal ujian yang

dikerjakan. Di samping itu skor mentah baru metupakan informasi awal yang berupa ringkasan data berupa angka dan memiliki makna kuantitatif yang lemah. Skor yang dihasilkan dari pengukuran akan lebih mudah diinterpretasikan serta memberikan makna yang luas terhadap hasil pengukuran, apabila disajikan dalam bentuk yang mudah difahami. Salah satu bentuk penyajian skor adalah dengan mengubah bentuk skor kuantitatif menjadi bentuk skor kualitatif yang berupa kategori tertentu.

Adanya faktor kekeliruan yang selalu muncul dalam pengukuran baik secara sistematis maupun acak, berakibat diperolehnya skor ujian yang tidak dapat menggambarkan kondisi responden yang sesungguhnya. Semakin besar tingkat kekeliruan, maka semakin jauh skor yang diperoleh dari skor yang sesungguhnya. Skor yang dimaksud adalah skor murni yang dapat menggambarkan kemampuan (*ability*) responden yang bersifat tidak tampak (laten). Skor yang didapat secara langsung dari hasil pengukuran adalah skor tampak (amatan) yang belum dapat menggambarkan skor sesungguhnya (skor murni). Untuk memdapatkan skor murni perlu dilakukan sebuah cara yang dinamakan dengan estimasi. Faktor kekeliruan dalam pengukuran juga berpengaruh pada munculnya disparitas (perbedaan) skor hasil ujian, baik pada skor tampak maupun pada skor murni. Disparitas skor dapat dilihat tidak hanya pada nilai skor yang dihasilkan, tetapi dapat juga dilihat berdasarkan kategori skor yang telah dibuat sebelumnya.

Pengukuran terhadap hasil belajar siswa membutuhkan pemahaman yang baik tentang cara menginterpretasikan skor hasil ujian agar menghasilkan informasi yang bermanfaat. Terdapat dua pendekatan yang dapat dilakukan dalam menginterpretasikan skor hasil ujian yaitu pendekatan teori klasik dan teori responsi butir (teori modern). Aminah (2013:1) menyebutkan bahwa pengukuran menurut teori klasik adalah pemberian angka kepada obyek atau kejadian dengan aturan tertentu, angka diartikan sebagai sifat yang melekat pada obyek. Menurut teori klasik karakteristik tes sangat dipengaruhi oleh kemampuan peserta tes (*ability*) yang menempuh tes tersebut. Jika peserta yang sama menempuh tes berbeda maka karakteristik peserta umumnya berubah. Jika peserta diberi tes yang mudah, kemampuan peserta dapat berada pada level tinggi. Sebaliknya jika diberikan tes yang sulit, kemampuan peserta dapat berada pada level rendah. Konsep yang terdapat dalam teori tes klasik pada dasarnya memiliki kemampuan untuk menjelaskan *error* yang terjadi pada pengukuran. Menurut Sarea (2019:1), model

error pengukuran ini sangat berkaitan dengan koefisien korelasi. Koefisien korelasi yang ditemukan oleh Charles Spearman merupakan upaya menjelaskan error menggunakan dua komponen yaitu korelasi skor murni dan korelasi skor tampak (amatan). Koefisien korelasi dalam teori tes klasik didasarkan pada teori bahwa nilai rata-rata pengukuran dari semua hasil pengukuran yang mungkin akan sama dengan pengukuran sebenarnya pada populasi. Beberapa parameter penting yang terdapat dalam teori tes klasik adalah error pengukuran yang bersifat acak dan pengukuran itu sendiri. Selanjutnya Sarea (2019:1) menjelaskan bahwa terdapat tiga komponen dalam pengukuran yakni: indikator yang tampak (amatan), indikator hipotetikal yang menunjukkan nilai populasi murni, dan konsep hipotetikal yang menunjukkan kuantitas ketidaksesuaian antara indikator sebenarnya dan indikator yang tampak (amatan).

Berbeda dengan teori klasik, menurut teori modern kemampuan seseorang tidak berubah karena karakteristik tes. Kedua teori ini digunakan sebagai pendekatan untuk mengukur kemampuan responden dengan tingkat kekeliruan sekecil mungkin. Setiap pendekatan memiliki asumsi dan teknik pengolahan data yang berbeda, namun memiliki keunggulan dan kelemahan masing-masing. Apabila dibandingkan antara penggunaan Teori klasik dan Modern secara praktis di bidang pendidikan dan psikologi, penggunaan teori klasik dalam pengukuran pendidikan masih dominan di sekolah-sekolah dasar dan menengah. Hal ini terkait dengan kepraktisan penggunaan teori klasik dan pemahaman guru-guru serta para praktisi pendidikan yang masih cenderung menggunakan teori klasik. Berdasarkan pada kondisi tersebut, maka pembahasan dalam penelitian ini cenderung menggunakan pendekatan teori klasik.

Ujian (tes) merupakan bentuk kegiatan pengukuran dalam bidang pendidikan atau psikologi (mental). Sebagai sebuah alat ukur yang baik, ujian seharusnya dilaksanakan sesuai dengan konsep dan teori pengukuran yang baik. Pelaksanaan suatu ujian harus disusun dengan konstruksi yang benar, dilaksanakan secara obyektif, dan dilakukan penskoran dan analisis yang tepat agar menghasilkan informasi yang benar dan bermanfaat untuk dijadikan bahan kajian serta perbaikan mutu pendidikan. Kadir (2015:1) menyebutkan bahwa tes yang baik harus memenuhi beberapa persyaratan, yaitu: harus efisien, harus baku, mempunyai norma, obyektif, valid (sahih), dan reliabel (andal). Kesimpulan dari penelitian yang dilakukan oleh Kadir adalah bahwa berdasarkan analisis

tes yang soal-soalnya valid, memiliki parameter tes yang berkategori baik dan baik sekali. Selanjutnya Nurjanah dan Marlianingsih (2015:1) menyimpulkan dari hasil penelitiannya tentang analisis butir soal Pilihan Ganda dari aspek kebahasaan bahwa analisis validitas bertujuan untuk mengkaji kesahihan alat ukur atau soal dalam menilai apa yang seharusnya diukur. Sedangkan analisis reliabilitas adalah untuk mengkaji keajegan (*stability*) atau ketetapan hasil tes, manakala tes tersebut diajukan kepada siswa yang sama lebih dari satu kali, atau dua perangkat tes yang setara kepada objek yang sama.

Alat ukur yang dikonstruksi dengan baik akan berpengaruh terhadap tingkat akurasi pengukuran. Sebuah alat ukur berupa perangkat soal ujian yang diberikan kepada responden akan menghasilkan skor yang akurat sehingga dapat menggambarkan kemampuan responden yang sebenarnya. Keakuratan alat ukur secara langsung akan mengurangi tingkat kekeliruan (*error*) dalam pengukuran. Skor yang akurat adalah skor tampak yang dihasilkan dari pengukuran yang nilainya mendekati atau hampir sama dengan skor sesungguhnya (skor murni). Untuk menghasilkan alat tes yang berkualitas, diperlukan alat ukur yang memiliki tingkat keajegan (reliabilitas) yang tinggi. Tingkat reliabilitas yang tinggi akan menghasilkan *error* yangkecil sehingga skor tampak yang dihasilkan melalui alat ukur tersebut semakin mendekati skor murni.

Secara teoritis penilaian hasil belajar siswa dapat dilakukan melalui dua pendekatan yaitu melalui Penilaian Acuan Patokan (PAP) dan Penilaian Acuan Normatif (PAN). Pengertian dari PAP adalah penilaian yang dilakukan dengan cara membandingkan hasil belajar setiap siswa dengan tingkat pencapaian kompetensi yang telah ditetapkan dalam tujuan pembelajaran. Sedangkan PAN merupakan pendekatan klasik yang membandingkan hasil belajar siswa pada suatu tes dengan hasil siswa lainnya pada tes yang sama. Menurut Suparman (2012), penilaian dengan menggunakan pendekatan acuan normatif (PAN) digunakan sebagai metode pengukuran yang menggunakan prinsip belajar kompetitif. Sedangkan penilaian dengan menggunakan pendekatan PAP merupakan pengukuran yang menggunakan acuan yang berbeda, yaitu hasil tes siswa dikomparasikan dengan kriteria lain yang telah ditentukan sebelumnya, bukan dengan hasil tes siswa lainnya. Dalam penelitian ini penulis melakukan analisis dengan menggunakan pendekatan PAN.

Penggunaan alat tes yang terukur dan terstandar oleh para pendidik di sekolah merupakan salah satu kunci berkualitasnya proses pembelajaran dan penilaian. Hasil belajar siswa merupakan indikator penting untuk menilai keberhasilan guru dalam pembelajaran. Skor siswa yang diperoleh dari tes yang dilakukan adalah parameter utama yang dapat digunakan untuk menilai tingkat kompetensi dan keberhasilan belajar siswa. Skor siswa adalah skor tampak yang diperoleh berdasarkan hasil tes. Secara teoritik, skor tampak sebenarnya belum bisa menggambarkan skor murni siswa. Secara ideal skor tampak seharusnya tidak berbeda jauh dengan skor murni yang dapat menggambarkan kemampuan siswa sesungguhnya, namun pada kenyataannya skor tampak yang dihasilkan dari sebuh tes seringkali tidak sesuai dengan kemampuan siswa yang sebenarnya. Masalah ini terjadi karena beberapa faktor yaitu sebaran skor tampak yang diperoleh dari hasil tes bisa sangat beragam dan berbeda secara signifikan dengan skor murninya. Di sisi lain kesenjangan antara skor tampak dengan skor murni dapat terjadi akibat kualitas (validitas dan reliabilitas) alat tes yang rendah. Perbedaan (disparitas) antara skor tampak dengan skor murni yang dipengaruhi oleh kualitas alat tes adalah ~~hal~~ fenomena yang menarik untuk diteliti. Pendekatan acuan normatif yang digunakan dalam melakukan analisis hasil tes akan memudahkan penulis untuk mengidentifikasi perbedaan skor dimaksud. Berdasarkan alasan inilah dilakukan kajian terhadap hasil belajar siswa ditinjau dari disparitas skor tampak dengan skor murni berdasarkan kategori acuan normatif.

Berdasarkan latar belakang yang telah diuraikan di atas, rumusan masalah dalam penelitian ini adalah sebagai berikut: 1) bagaimana estimasi terhadap skor murni (T) berdasarkan skor tampak (X) dari hasil belajar siswa? 2) bagaimana pengkategorian skor tampak (X) dan skor murni (T) berdasarkan acuan normatif pada hasil belajar siswa? 3) bagaimana disparitas hasil pengkategorian skor tampak (X) dan skor murni (T) pada hasil belajar siswa?

B. Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah metode kuantitatif .Sumber data yang digunakan dalam penelitian ini adalah respon peserta ujian nasional SMP pada mata pelajaran matematika tahun 2016/2017 sebanyak 40 butir dengan jumlah responden

sebanyak 250 orang. Selanjutnya jumlah responden yang terpilih tersebut dikelompokkan menjadi 5 kelompok (grup) yang masing-masing berjumlah 50 responden.

Pengolahan dan analisis data yang dilakukan dalam penelitian ini pada masing-masing grup sebagai berikut:

1. Menghitung rerata skor tampak (X);
2. Menghitung koefisien reliabitas tes Alpha Crocbach dengan menggunakan program SPSS versi 19;
3. Melakukan estimasi skor murni (T) berdasarkan tiap-tiap skor tampak (X) dengan menggunakan rumus (4);
4. Membuat batas-batas kategori acuan normatif berdasarkan skor tampak (X) dengan menggunakan rumus (2) dan (3);
5. Mengkategorikan dengan acuan normatif tiap-tiap skor tampak (X) dan skor murni (T) berdasarkan batas-batas kategori yang diperoleh pada langkah 4;
6. Melakukan analisis disparitas skor tampak (X) dan skor murni (T) bertdasarkan kategori acuan normatif.

Skor yang diperoleh dari hasil pengukuran secara umum dapat menggambarkan performansi responden. Menurut Azwar (2015:68), skor tidak lain daripada harga suatu jawaban terhadap pertanyaan dalam ujian yang - meskipun tidak sempurna - merupakan representasi dari suatu atribut laten. Yang dimaksud dengan atribut laten, Naga (2013:13) menjelaskan bahwa atribut laten merupakan atribut yang tidak dapat langsung diukur. Hasil belajar merupakan contoh atribut laten karena tidak dapat diukur secara langsung dan pengukurannya dapat dilakukan dengan cara mencari atribut manifes (atribut yang dapat langsung diukur) yang sepadan dengan atribut laten yang akan diukur. Selanjutnya Naga (2013:13) menjelaskan dalam dunia pendidikan dan psikologi yang selama ini berkembang diyakini bahwa sekor hasil ujian adalah salah satu bentuk atribut manifes yang sepadan dengan keberhasilan belajar siswa yang merupakan atribut laten.

Hasil pengukuran ujian yang dinyatakan dengan skor menggambarkan ukuran kuantitatif dari pengukuran tersebut. Skor kuantitatif ini dapat secara langsung diperoleh setelah dilakukan penskoran. Azwar (2013:68) menjelaskan bahwa skor tersebut merupakan skor perolehan (*obtained scores* atau *observed scores*) yang belum diolah atau diderivasikan yang selanjutnya disebut dengan skor tampak atau skor amatan dan diberi

simbol X. Bersamaan dengan itu, dari hasil pengukuran sekaligus akan diperoleh skor lainnya yang merupakan skor sesungguhnya atau skor murni atau skor tulen (*true score*) yang menggambarkan performansi yang sebenarnya dan merupakan representasi murni dari atribut laten. Skor murni (skor tulen) tersebut biasanya dilambangkan dengan simbol T. Selanjutnya menyertai setiap hasil pengukuran akan selalu diperoleh kekeliruan (*error*) pengukuran yang besarnya tidak diketahui secara langsung. Skor tersebut dinamakan skor keliru (*error*) dan biasanya dilambangkan dengan simbol E.Skor tampak (X) dapat diketahui secara langsung dari hasil pengukuran, sedangkan skor murni (T) dan skor keliru (E) tidak dapat langsung diketahui karena bersifat tersembunyi (laten).

Untuk mendapatkan skor murni (T) dan skor keliru (E) terlebih dahulu harus dilakukan sebuah cara yang dinamakan dengan estimasi. Terdapat beberapa metode atau pendekatan yang dapat dilakukan untuk mengestimasi skor. Masing-masing metode/pendekatan estimasi memiliki prosedur dan keunggulan tertentu, sehingga dalam pelaksanaannya perlu dipilih metode yang sesuai yang akan dilakukan. Berdasarkan hasil penelitian yang dilakukan oleh Widayati (2009) tentang komparasi beberapa metode estimasi kesalahan pengukuran, diperoleh kesimpulan bahwa estimasi skor keliru (*error*) dapat dilakukan dengan beberapa metode/pendekatan teori klasik dan teori responsi butir. Dari hasil penelitiannya disebutkan bahwa secara umum perbedaan hasil estimasi dari metode-metode atau pendekatan tersebut menghasilkan nilai yang tidak terlalu jauh. Dengan kata lain, pendekatan estimasi skor murni yang digunakan tersebut memiliki efektivitas yang hampir sama. Pada bagian lain, Widhiarso (2018) dalam hasil simulasinya tentang estimasi terhadap skor murni menyebutkan sebuah kesimpulan bahwa estimasi skor murni akan mendekat pada rerata dibandingkan dengan skor tampak. Hal ini mengandung arti bahwa hasil estimasi skor murni akan lebih akurat dibandingkan dengan skor tampak.

Secara konseptual sejak tahun 1910 Spearman telah menurumuskan hubungan antara skor tampak (X), skor murni (T), dan skor keliru (E) dalam sebuah persamaan matematis sebagaimana McDonald (1999:64) dan Nitko (1983:389) menggambarkan persamaan tersebut sebagai berikut:

true score = obtained score - error score (T = X - E) atau
obtained score = true score + error score (X = T + E) …(1)

Skor tampak (skor amatan) merupakan jumlah dari dua skor yaitu skor murni (skor tulen) dan skor keliru (*error*). Sedangkan skor keliru (*error*) memperlihatkan besarnya kekeliruan dalam pengukuran. Nilai *error* bisa berharga positif, negatif, atau nol. *Error* juga dapat menyebabkan skor tampak lebih tinggi atau lebih rendah daripada skor murni.

Hubungan yang berlaku berdasarkan persamaan (1) adalah bersifat aditif. Azwar (2013:69) menjelaskan apabila tidak terjadi *error* dalam pengukuran (E=0), maka skor tampak (X) akan identik dengan skor murni (T). Sebaliknya apabila dalam pengukuran terjadi *error*, maka akan muncul *error* negatif yang menyebabkan skor tampak (X) akan lebih rendah nilainya dari skor murni (T), dengan kata lain akan muncul underestimasi skor tampak (X) terhadap skor murni (T). Selanjutnya apabila diperoleh *error* positif, maka akan menghasilkan skor tampak (X) melebihi skor murni (T), dengan kata lain akan muncul overestimasi skor tampak (X) terhadap skor murni (T).

Kekeliruan pengukuran dapat terjadi secara sistematis (*systemartic error*) atau terjadi secara acak (*random errror*). Widhiarso (2018) menjelaskan kekeliruan sistematis dapat disebabkan oleh kekeliruan atau kesalahan dalam mempersiapkan alat ukur atau menyelenggarakan ujian. Senada dengan Widhiarso (2018), Widayati (2009) menjelaskan bahwa kesalahan yang bersifat sistematis disebabkan oleh orang yang mengukur atau alat ukur/instrumennya. Kesalahan pengukuran yang sistematis adalah kesalahan yang secara konsisten mempengaruhi hasil pengukuran. Gangguan suara gaduh lingkungan pada saat ujian dapat menurunkan hasil ujian yang diikuti oleh siswa atau penyusunan pertanyaan-pertanyaan dalam butir soal ujian yang hanya menguntungkan beberapa pihak dapat mengurangi obyektifitas hasil ujian. Kekeliruan sistematis ini dapat disebut juga dengan bias dan secara matematis akan mempengaruhi hasil pengukuran secara konsisten. Jenis kekeliruan lainnya adalah kekeliruan acak. Widhiarso (2018) menjelaskan bahwa kekeliruan acak dapat disebabkan oleh faktor-faktor acak yang tidak dapat diperkirakan sebelumnya yang dapat mempengaruhi hasil pengukuran namun secara tidak konsisten. Sedangkankan Widayati (2009) menyebutkan bahwa kesalahan bersifat acak tidak memiliki pola secara sistematis. Kesalahan acak disebabkan antara lain karena kesalahan dalam menentukan sampel isi tes, dan adanya variasi emosi seseorang bersifat acak.

Pemahaman terhadap interpretasi skor dapat lebih mudah dilakukan dengan cara mengubah skor kuantitatif ke dalam skor kualitatif. Biasanya skor yang diperoleh dari hasil ujian belum langsung dapat diterjemahkan secara mudah. Skor tersebut dapat dikonversi ke dalam bentuk kategori sehingga kedudukan skor tersebut dapat dibandingkan dengan kedudukan skor-skor lainnya. Azwar (2015:145) menjelaskan bahwa pada dasarnya interpretasi skor pada skala psikologi berisfat normatif, artinya makna skor diartikan pada posisi relatif skor terhadap suatu norma (mean) skor populasi teoritik sebagai parameter sehingga hasil ukur yang berupa angka (kuantitatif) dapat diinterpretasikan secara kualitatif. Dalam hal ini acuan normatif dapat memudahkan dalam memahami hasil pengukuran. Terdapat beberapa pendekatan dalam membuat kategorisasi, salah satunya adalah cara kategorisasi dengan pertimbangan eror standar dalam pengukuran.

Kecermatan dalam melakukan pengukuran tidak hanya dapat dilakukan dengan mempertimbangkan koefisien reliabiltas tapi juga diperlukan pertimbangan lainnya yaitu dengan memperhitungkan eror standar dalam pengukuran (*standard error of measurement* - *SEM*) . Crocker dan Aigina (2008:122-123) menuliskan rumus SEM sebagai berikut:

$$\sigma_E = \sigma_X \sqrt{1-\rho xx'} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

dimana:

σ_E = standar eror

σ_X = standar deviasi skor

$\rho xx'$ = koefisien reliabilitas tes

Selanjutnya prosedur yang perlu dilakukan untuk menentukan batas-batas kategori skor adalah menghitung interval kepercayaan skor murni yang diestimasi. Rumus untuk menentukan interval kepercayaan skor murni sebagaimana dituliskan oleh Azwar (2015:80) adalah sebagai berikut:

$$X - (Z_{\alpha/2})\sigma_E \leq T \leq X + (Z_{\alpha/2})\sigma_E \quad \ldots\ldots..(3)$$

dimana:

T = skor murni yang diestimasi

X = skor tampak

$Z_{\alpha/2}$ = harga Z pada taraf signifikansi $\sigma/2$

σ_E = standar error

Untuk mengestimasi skor murni digunakan rumus estimasi skor murni yang diperoleh dari rumus regresi linier sebagaimana dituliskan oleh Crocker dan Aigina (2008:147) sebagai berikut:

$$T = \rho xx'(X - \mu_x) + \mu_x \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(4)$$

dimana:

T = skor murni yang diestimasi

X = skor tampak

$\rho xx'$ = koefisien reliabilitas tes

μ_X = rerata skor

C. Hasil Penelitian dan Pembahasan

Proses untuk menghasilkan estimasi skor murni dari sebuah hasil ujian dengan menggunakan pendekatan teori tes klasik merupakan rangkaian perhitungan dan pengolahan data yang memerlukan beberapa data statistik. Sesuai dengan rumusan estimasi skor murni yang dituliskan oleh Crocker dan Aigina (2008:147), data statistik skor tampak yang dibutuhkan untuk mengestimasi skor murni adalah rata-rata (mean), standar deviasi, varians, koefisien reliabilitas tes, standar error, batas bawah nilai, dan batas atas nilai. Data statistik tersebut dihitung pada masing-masing kelompok (grup) responden. Berikut adalah Tabel 1 yang berisi data statistik yang telah dihitung dengan program SPSS pada masing-masing grup.

Tabel 1
Statistik Skor Tampak (X) pada Tiap-tiap Grup

Statistik	Grup1	Grup2	Grup3	Grup4	Grup5
Mean μ_x	20,080	20,360	22,180	17,740	17,680
Stand.Deviasi σ_X	8,836	9,454	10,978	6,486	6,900
Variansi σ^2_X	78,075	89,378	120,518	42,074	47,610
Koef.Reliabilitas Tes $\rho xx'$	0,897	0,912	0,942	0,788	0,815
SEM σ_E	2,836	2,805	2,644	2,987	2,968
Batas Bawah Kategori	15,401	15,733	17,818	12,812	12,783
Batas Atas Kategori	24,759	24,987	26,542	22,668	22,577

Berdasarkan hasil perhitungan statistik dengan menggunakan rumus dan program SPSS versi 19 pada Tabel 1 tampak bahwa koefisien reliabilitas tes **Alpha Cronbach** pada tiap-tiap grup bervariasi. Koefisien reliabilitas tertinggi terdapat sebesar 0,942

terdapat pada grup 3 dan koefisien reliabilitas terendah sebesar 0,788 terdapat pada grup 4. Secara umum dari keseluruhan koefisien reliabilitas tes tampak bahwa nilainya masing-masing lebih dari 0,7. Nilai koefisien reliabilitas tersebut menurut Pallant (2016:101) dikategorikan sebagai nilai koefisien yang ideal. Hal ini menunjukkan internal konsistensi yang tinggi dari alat ukur yang digunakan. Sehingga alat ukur yang digunakan tersebut ajeg (reliabel) dan layak untuk digunakan sebagai instrumen.

Besar kecilnya koefisien reliabilitas berbanding terbalik dengan standar eror pengukuran (SEM). Pada grup 3 nilai SEM paling rendah yaitu sebesar 2,644 dan nilai SEM tertinggi terdapat pada grup 4 yaitu sebesar 2,987. Semakin tinggi nilai SEM menunjukkan semakin besar kekeliruan yang terjadi pada pengukuran dan kekeliruan yang besar tersebut muncul disebabkan oleh alat ukur yang kurang reliabel. Sebaliknya nilai SEM yang rendah menunjukkan semakin kecilnya kekeliruan dalam pengukuran dan sekaligus menunjukkan bahwa alat ukur yang digunakan lebih reliabel.

Berdasarkan hasil perhitungan dan analisis yang dilakukan sesuai dengan Tabel 1, informasi yang dapat diperoleh adalah alat tes yang digunakan pada siswa Grup 3 merupakan alat tes yang paling reliabel (handal). Hal ini menunjukkan tingkat kepercayaan hasil tes yang paling tinggi dibandingkan dengan hasil tes pada Grup lainnya. Sebaliknya alat tes yang digunakan pada siswa Grup 4 merupakan alat tes yang paling rendah kehandalannya. Tingkat kualitas alat tes yang paling handal pada Grup 3 juga ditunjukkan dengan nilai error (SEM) yang paling rendah. Demikian sebaliknya nilai error (SEM) yang paling tiggi pada alat tes Grup 4 menunjukkan kualitas alat tes yang paling rendah.Safari (2020:294) menjelaskan bahwa standar error pengukuran (SEM) berguna untuk mengetahui besarnya faktor kesalahan pengukuran suatu tes. Semakin kecil nilai SEM suatu tes, maka semakin konsisten skor-skor suatu tes. Nilai SEM selalu berlawanan dengan koefisien reliabilitas. Jika reliabilitas suatu tes tinggi, maka nilai SEM nya rendah atau sebaliknya.

Tahap berikutnya adalah melakukan estimasi terhadap skor murni (T) berdasarkan skor tampak (X) pada setiap grup dengan menggunakan rumus (4). Hasil estimasi skor murni (T) menunjukkan bahwa skor murni hasil estimasi (T) nilainya lebih mendekati rerata (mean) skor tampak (X). Menurut Azwar (2015:84) dikarenakan estimasi terhadap reliabilitas hasil ukur yang dinyatakan dalam bentuk koefisien reliabilitas yang dihitung

berdasarkan pada sampel data empiris akan selalu lebih kecil dari 1,00 maka estimasi terhadap skor murni (T) deviasinya akan selalu lebih kecil terhadap rerata dibandingkan skor tampak (X). Berdasarkan hasil perhitungan, estimasi terhadap skor murni (T) pada setiap individu akan selalu menghasilkan angka yang lebih dekat dengan rerata (mean) skor tampak (X) kelompok daripada skor tampak (X) itu sendiri. Semakin besar nilai koefisien reliabilitasnya maka akan semakin dekat hasil estimasi dengan rerata skor.

Langkah selanjutnya adalah melakukan pengkategorian skor tampak (X) dan skor murni (T) berdasarkan acuan normatif pada masing-masing grup. Kategori yang digunakan terdiri dari kategori skor rendah, sedang, dan tinggi. Untuk menentukan panjang interval masing-masing kategori, terlebih dahulu dihitung batas bawah dan batas atas kategori dengan menggunakan rumus (2) dan (3) dan diperoleh nilai-nilai batas bawah dan batas atas sebagaimana tercantum dalam Tabel 1. Adanya nilai batas bawah dan nilai batas atas yang telah dihitung, menjadikan terdapatnya dua titik batas sehingga muncul tiga buah kategori (rendah, sedang, tinggi). Nilai batas atas dan batas bawah menurut Mardapi (2000:9) disebut taksiran interval yaitu dua bilangan numerik yang ditentukan berdasarkan data sampel dan diikuti dengan pernyataan besarnya suatu interval kepercayaan bahwa suau interval mencakup besarnya parameter yang diukur. Penentuan kategori ini juga dengan mempertimbangkan standar eror pengukuran (SEM). Alasan digunakannya SEM dalam menentukan kategori ini adalah untuk memperhitungkan tingkat kecermatan pengukuran. Rentang nilai yang dijadikan dasar pengkategorian nilai hasil estimasi skor murni ditunjukkan pada Tabel 2.

Tabel 2
Rentang Nilai Hasil Estimasi Skor Murni (T) Berdasarkan Skor Tampak (X)
Jumlah Responden Tiap Grup (M)=50, Jumlah Butir Soal Tiap Grup (N)=40

Grup	Rentang Nilai Rendah	Rentang Nilai Sedang	Rentang Nilai Tinggi
1	7,0 – 15,3	15,4 – 24,7	24,8 – 40,0
2	8,0 – 15,6	15,7 – 24,8	24,9 – 40,0
3	8,0 – 17,7	17,8 – 26,4	26,5 – 40,0
4	6,0 – 12,7	12,8 – 22,6	22,7 – 40,0
5	7,0 – 12,6	12,7 – 22,5	22,6 – 40,0

Pada Tabel 2 diperoleh informasi rentang nilai untuk masing-masing kategori (rendah, sedang, tinggi). Rentang nilai tersebut berbeda pada tiap-tiap grup tergantung pada nilai batas bawah dan batas atas yang sebelumnya sudah dihitung sebagaimana terdapat pada Tabel 1. Rentang nilai masing-masing kategori pada tiap-tiap grup akan menjadi pedoman untuk

menghitung banyaknya skor yang termasuk pada masing-masing kategori-kategori tersebut. Setelah semua skor dimasukkan ke dalam tiap-tiap kategori, selanjutnya akan diketahui perbedaan penyebaran skor tiap-tiap kategori pada masing-masing grup. Perbedaan inilah yang selanjutnya dijadikan pedoman dalam mengukur disparitas skor.

Langkah berikutnya adalah memasukan skor tampak (X) dan skor murni (T) ke dalam kategori-kategori yang sudah ditentukan sebagaimana terdapat pada Tabel 2. Masing-masing kategori pada setiap grup berisi angka yang menunjukkan banyaknya skor tampak (X) dan skor murni (T). Hasil pengelompokkan skor-skor tersebut dapat dilihat pada Tabel 3.

Tabel 3
Hasil Pengkategorian Skor pada Tiap-tiap Grup

Grup	Kategori Skor	Banyaknya Skor Rendah	Banyaknya Skor Sedang	Banyaknya Skor Tinggi
Grup1	Skor Tampak (X)	18	19	13
	Skor Murni (T)	14	23	13
Grup2	Skor Tampak (X)	21	16	13
	Skor Murni (T)	21	18	11
Grup3	Skor Tampak (X)	23	8	19
	Skor Murni (T)	23	10	17
Grup4	Skor Tampak (X)	11	28	11
	Skor Murni (T)	7	36	7
Grup5	Skor Tampak (X)	14	22	14
	Skor Murni (T)	11	27	12

Tabel 3 memperlihatkan banyaknya skor tampak (X) dan estimasi skor murni (T) yang telah dikelompokkan ke dalam tiga kategori yaitu: rendah, sedang, dan tinggi. Kategori-kategori tersebut merupakan kategori acuan normatif karena didasarkan pada rerata (mean) pada masing-masing grup secara terpisah. Berdasarkan hasil pengelompokkan inilah diperoleh perbedaan banyaknya skor pada masing-masing kategori dan perbedaan inilah yang memunculkan disparitas skor tampak (X) dan skor murni (T). Pengkategorian acuan normatif pada tiap-tiap Grup didasarkan pada hasil perhitungan nilai batas bawah dan nilai batas atas sebagaimana terdapat pada Tabel 1. Berdasarkan data pada Tabel 3 diperoleh informasi bahwa pada masing-masing Grup terdapat perbedaan (disparitas) antara skor tampak (X) dengan skor murni (T) di setiap kelompok kategori. Disparitas sebagaimana dimaksud dapat dilihat dari banyaknya skor tampak (X) dan skor murni (T) setiap kategori (rendah, sedang, tinggi) yang berbeda-

beda pada masing-masing Grup. Besarnya perbedaan pada masing-masing Grup ini yang selanjutnya dijadikan ukuran tingkat disparitasnya.

Langkah akhir yang dilakukan dalam penelitian ini adalah melakukan analisis disparitas skor tampak (X) dan skor murni (T) pada tiap-tiap grup dengan memperhatikan koefisien reliabilitas tes pada masing-masing grup. Hasil analisis disparitas skor-skor tersebut dapat dilihat pada Tabel 4.

Tabel 4
Disparitas Skor Tampak (X) dan Skor Murni (T) Berdasarkan Kategori

Grup	$\rho xx'$	Δ_{X-T}
Grup 4	0,788	16
Grup 5	0,815	10
Grup 1	0,897	8
Grup 2	0,912	4
Grup 3	0,942	4

Keterangan:

$\rho xx'$ = Koefisien reliabilitas tes

Δ_{X-T} = Disparitas skor berdasarkan kategori antara X dan T

Berdasarkan Tabel 4 dapat diketahui bahwa terdapat disparitas skor antara skor tampak (X) dengan skor murni (T) berdasarkan kategori yang telah ditentukan. Adanya disparitas ini dipengaruhi oleh besarnya koefisien reliabilitas tes. Disparitas skor yang paling tinggi terdapat pada grup 4 yaitu 16, artinya pada Grup 4 terdapat perbedaan skor tampak (X) dengan skor murni (T) sebanyak 16 skor untuk semua kategori acuan normatif (rendah, sedang, tinggi). Disparitas yang paling kecil terdapat pada grup 2 dan grup 3 yaitu masing-masing 4, artinya pada Grup 2 terdapat perbedaan skor tampak (X) dengan skor murni (T) sebanyak 4 skor untuk semua kategori acuan normatif (rendah, sedang, tinggi).

Apabila dilihat dari banyaknya perbedaan skor antara Grup 2 dan Grup 3, nampaknya kedua grup tersebut memiliki tingkat disparitas yang sama. Selanjutnya untuk membedakan tingkat disparitas masing-masing grup selain berdasarkan banyaknya perbedaan skor, dapat pula dilihat berdasarkan nilai reliabilitasnya. Grup 3 memiliki nilai reliabilitas yang lebih tinggi (0,942) daripada nilai reliabilitas Grup 2 (0,912). Secara teoritik, semakin kecil koefisien reliabilitas tes, maka semakin tinggi tingkat disparitasnya,

dan sebaliknya semakin besar koefisien reliabilitas tes, maka semakin rendah tingkat disparitasnya. Ukuran tingkat disparitas suatu hasil pengukuran menunjukkan adanya tingkat akurasi dan sekaligus menentukan kualitas hasil pengukurannya. Tingkat akurasi dan ketepatan hasil pengukuran tidak terlepas dari koefisien reliabilitas alat ukurnya. Arikunto (2010) menyebutkan bahwa tingkat reliabilitas menunjukkan sejauhmana suatu pengukuran dapat dipercaya karena keajegannya. Hal ini menunjukkan bahwa jika alat ukur yang digunakan memiliki kualitas yang tinggi (ditunjukkan dengan koefisien reliabilitasnya yang tinggi), maka kekeliruan dalam pengukuran akan semakin kecil sehingga hasil pengukuran skor tampak (X) akan semakin mendekati skor murni (T). Dengan demikian alat tes pada Grup 3 memiliki tingkat disparitas yang paling rendah, artinya alat tes pada Grup 2 dapat dikategorikan sebagai alat tes yang memiliki kualitas yang paling baik.

D. Kesimpulan

Berdasarkan pada analisis data yang telah dilakukan, maka simpulan dari penelitian ini sebagai berikut.1) skor murni (T) berdasarkan skor tampak (X) merupakan pendekatan yang tepat untuk mengetahui skor dari hasil belajar siswa yang mendekati nilai sebenarnya. 2) Pengkategorian skor hasil belajar siswa yang dilakukan berdasarkan acuan normatif berfungsi untuk membantu guru dalam mengetahui disparitas tingkat kemampuan siswa berdasarkan estimasi skor murni (T) terhadap skor tampak (X). 3) Disparitas skor hasil estimasi dipengaruhi oleh tingkat reliabltas alat tes. Semakin tinggi tingkat reliabilitas alat tes, maka semakin rendah disparitasnya yang berarti semakin baik alat tes tersebut untuk mengukur kemampuan siswa. Kesimpulannya adalah untuk menghasilkan disparitas skor yang rendah dan memperoleh data yang akurat, maka alat tes yang digunakan harus memiliki tingkat reliabilitas yang tinggi.

Daftar Pustaka

Aminah, Nonoh Siti. 2013. Teori Pengukuran Dalam Pendidikan. *Jurnal Materi dan Pembelajaran Fisika (JMPF),*Vol(3) No 2, 33-39.

Arikunto, Suharsimi. 2010. Prosedur penelitian : suatu pendekatan praktik. Jakarta : Rineka Cipta.

Azwar, Saifuddin. 2015. Dasar-dasar Psikometrika. Yogyakarta: Pustaka Pelajar.

_____. 2015. *Penyusunan Skala Psikologi*. Yogyakarta: Pustaka Pelajar.

Crocker, Linda & James Aigina. 2008. *Introduction to Classical and Modern Test Theory*. Ohio: Cengange Learning.

Kadir, Abdul. 2015. Menyusun dan Menganalisis Tes Hasil Belajar. *Jurnal Al-Ta'dib*, Vol (8) No.2, 70-81.

Mardapi, Djemari. 2000. *Pengujian Hipotesis Nihil: Uji Signifikansi dan Interval Kepercayaan*. Yogyakarta: Buletin Psikologi.

McDonald, Roderick P. 1999. *Test Theory: A Unfied Treatment.* New Jersey: Lawrence Eribaum Associates.

Naga, Dali Santun. 2013. *Teori Sekor pada Pengukuran Mental*. Jakarta: PT Nagarani Citrayasa.

_____1992. *Pengantar Teori Sekor pada Pengukuran Pendidikan*. Jakarta: Gunadarma.

Nitko, Anthony J. 1983. *Educational Test and Measurment An Introduction*. New York: Harcout Brace Jovanovich.

Nurjanah & Noni Marlianingsih. 2015. Analisis Butir Soal Pilihan Ganda dari Aspek Kebahasaan. *Faktor Jurnal Ilmu Kependidikan*, Vol (II) No. 1, 69-78.

Pallant, Julie. 2016. *SPSS Survival Manual*. Sydney-Melbourne-Auckland-London: Allen & Unwin.

Safari. 2020. *Statistika untuk Penelitian Bahasa, Bimbingan Konseling, Psikologi, Hukum, Agama, Teknik, Ekonomi, Keperawatan, Kedokteran, Paud, dan Pendidikan*. Jakarta: Universitas Islam As-Syafi'iyah.

Sarea, Muh.Syahrul dan Rosnia Ruslan. 2019. Karakteristik Butir Soal: Classical Test Theory Vs Item Response Theory. *Didaktika Jurnal Kependidikan*, Vol (13) No 1, 1-16

Sumintono, Bambang & Wahyu Whidiarso. 2015. *Aplikasi Pemodelan Rasch: pada Assessment Pendidikan*. Cimahi: Trim Komunikata.

Suparman, M.A. 2012. *Desain Instruksional Modern*. Jakarta: Erlangga.

Widayati, Catharina Sri Wahyu. 2009. Komparasi Beberapa Metode Estimasi Kesalahan Pengukuran. *Jurnal Penelitian dan Evaluasi Pendidikan*, 12 (2), 182 - 197.

http://widhiarso.staff.ugm.ac.id/files/Eror%20Pengukuran.pdf