

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB II

TINJAUAN PUSTAKA

A. Karakteristik Psikometri

Karakteristik psikometri merupakan atribut kuantitatif yang berhubungan dengan kekuatan atau kelemahan statistik dari tes atau pengukuran (Embretson & Reise, 2000). Karakteristik psikometri mencakup reliabilitas, validitas, tingkat kesukaran aitem, dan daya diskriminasi aitem (Embretson & Reise, 2000). Estimasi karakteristik psikometri pada penelitian ini menggunakan pendekatan *Rasch Model*, yang meliputi unidimensionalitas, reliabilitas, *item fit order*, indeks kesukaran aitem, dan *differential item functioning* (DIF).

1. Unidimensionalitas

Asumsi umum dari model IRT adalah bahwa hanya satu kemampuan diukur dengan satu set aitem dalam tes, hal inilah yang disebut dengan *unidimensionality* (Hambleton, dkk., 1991; Hambleton & Swaminathan, 1985). Asumsi ini tidak dapat dipenuhi secara ketat karena faktor kognitif, kepribadian, dan keikutsertaan dalam tes selalu mempengaruhi hasil tes, setidaknya sampai batas tertentu (Hambleton, dkk., 1991; Hambleton & Swaminathan, 1985). Faktor-faktor ini termasuk tingkat motivasi, kecemasan tes, kemampuan untuk bekerja dengan cepat, kecenderungan untuk menebak bila ragu tentang jawaban, dan keterampilan kognitif (Hambleton & Swaminathan, 1985). Satu set data tes perlu adanya komponen "dominan" atau faktor yang mempengaruhi hasil tes, agar asumsi ini dapat terpenuhi (Hambleton, dkk., 1991).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Unidimensionalitas instrumen merupakan ukuran yang penting untuk mengevaluasi apakah instrumen yang dikembangkan mampu mengukur apa yang seharusnya diukur (Sumintono & Widhiarso, 2014). Persyaratan unidimensionalitas minimal sebesar 20% dapat terpenuhi, apabila nilainya lebih dari 40% artinya lebih bagus, jika nilainya 60% artinya istimewa (Sumintono & Widhiarso, 2014). Nilai varians yang tidak dapat dijelaskan oleh instrumen idealnya tidak melebihi 15% (Sumintono & Widhiarso, 2014).

2. Reliabilitas

Reliabilitas merupakan terjemahan dari kata *reliability* yang berasal dari kata *rely* dan *ability* (Azwar, 2009, 2014a). Sumintono dan Widhiarso (2014) memaparkan bahwa reliabilitas menunjukkan seberapa jauh pengukuran yang dilakukan berulang kali akan menghasilkan informasi yang sama, artinya tidak menghasilkan banyak perbedaan informasi yang berarti. Perbedaan informasi akan selalu ada, oleh karena itu, pengukuran yang meyakinkan tidak harus menghasilkan informasi yang sama persis, tetapi perbedaan-perbedaan itu nilainya kecil dan masih dapat diterima selagi berada dalam batas toleransi (Azwar, 2009, 2014a; Sumintono dan Widhiarso, 2014).

Reliabilitas skor tes secara umum dapat dilihat melalui perhitungan *Cronbach's alpha* dan KR-20, berkisar dari 0 – 1 (Boone, Staver, Yale, 2014). Nilai 0 menunjukkan benar-benar ketidak-reliabilitas pengukuran, dan nilai 1 menunjukkan reliabilitas sempurna (Aiken & Marnat, 2008; Azwar, 2009). Nilai KR-20 dikategorikan sebagai berikut: < 0,5 tergolong buruk sekali, 0,5-0,6 tergolong



buruk, 0,6-0,7 tergolong cukup, 0,7-0,8 tergolong bagus, dan $> 0,8$ tergolong bagus sekali (Sumintono & Widhiarso, 2014, 2015). Walaupun koefisien korelasi dapat saja bertanda negatif (-), koefisien reliabilitas selalu mengacu pada angka positif (+), karena angka yang negatif tidak ada artinya dalam interaksi reliabilitas hasil ukur (Azwar, 2009).

Pada pengukuran *Rasch Model* terdapat dua indeks reliabilitas yaitu *person reliability* dan *item reliability* (Boone, Staver, Yale, 2014). Nilai *person reliability* dan *item reliability* yakni apabila $< 0,67$ tergolong lemah, 0,67-0,80 tergolong cukup, 0,81-0,90 tergolong bagus, 0,91-0,94 tergolong bagus sekali, dan $> 0,94$ tergolong istimewa (Sumintono & Widhiarso, 2014, 2015). Konsistensi yang sempurna, yakni apabila nilai reliabilitas adalah 1, hal itu tidak dapat terjadi pada pengukuran aspek-aspek psikologi dan sosial yang menggunakan manusia sebagai subjeknya, karena terdapatnya berbagai sumber eror dalam diri manusia (Azwar, 2009).

3. Item Fit Order

Item fit menjelaskan apakah butir soal berfungsi normal melakukan pengukuran atau tidak (Sumintono & Widhiarso, 2015). Sumintono dan Widhiarso (2014) menjelaskan tentang seberapa akurat data sesuai dengan model ideal berkaitan dengan statistik kesesuaian, yang terdiri dari *infit*, *outfit*, *mean-square*, dan *standardized fit statistic* (ZSTD). *Infit* (*inlier-sensitive fit*) merupakan kesensitifan pola respon terhadap aitem sasaran pada responden atau sebaliknya (Sumintono & Widhiarso, 2014). *Outfit* (*outlier-sensitive fit*) mengukur

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

kesensitifan pola respon terhadap aitem dengan tingkat kesulitan tertentu pada responden atau sebaliknya (Sumintono & Widhiarso, 2014).

Aitem yang tidak *fit* mengindikasikan bahwa terjadi miskonsepsi pada responden terhadap butir soal (Sumintono & Widhiarso, 2015). Memeriksa aitem yang *fit* dan *misfit* dapat menggunakan nilai INFIT MNSQ dari setiap aitem; mean dan standar deviasi dijumlahkan, kemudian dibandingkan, nilai logit yang lebih besar menunjukkan aitem yang *misfit* (Sumintono & Widhiarso, 2014). Kriteria lain yang dapat digunakan adalah (Sumintono & Widhiarso, 2014; 2015):

- 1) Nilai *Outfit Mean Square* (MNSQ) yang diterima: $0,5 < \text{MNSQ} < 1,5$
- 2) Nilai *Outfit Z-Standard* (ZSTD) yang diterima: $-2,0 < \text{ZSTD} < +2,0$
- 3) Nilai *Point Measure Correlation* (*Pt Measure Corr*): $0,4 < \text{Pt Measure Corr} < 0,85$

Mean-square fit statistic menunjukkan jumlah distorsi dalam sistem pengukuran (Sumintono & Widhiarso, 2014). Jika nilai kurang dari 0,5 itu artinya aitem mudah ditebak, sedangkan nilai yang lebih besar dari 1,5 berarti aitem tidak mudah diprediksi (Sumintono & Widhiarso, 2014). *Standardized fit statistic* (ZSTD) berfungsi untuk melihat kesesuaian data dengan model (Sumintono & Widhiarso, 2014). Hasilnya adalah nilai z yang menjelaskan signifikansi jika data memang sesuai dengan model (Sumintono & Widhiarso, 2014).

Apabila aitem tidak memenuhi kriteria nilai *outfit* MNSQ, *outfit* ZSTD, dan *point measure correlation*, dapat dipastikan bahwa aitem kurang bagus sehingga perlu diperbaiki atau diganti (Sumintono & Widhiarso, 2015). Sumintono dan

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Widhiarso (2015) juga menjelaskan bahwa kesalahan menetapkan kunci jawaban, banyaknya individu yang kurang termotivasi dalam mengerjakan soal, dan soal yang memiliki daya beda rendah dapat menurunkan ketepatan butir.

4. Parameter Kesukaran Aitem (b)

Parameter kesukaran aitem adalah parameter lokasi, yang menunjukkan posisi *item characteristic curve* (ICC) dalam kaitannya dengan skala kemampuan (Hambleton, Swaminathan, Rogers, 1991). Parameter kesukaran aitem memiliki probabilitas respons yang benar sebesar 50% (Hambleton, Swaminathan, Rogers, 1991). Semakin besar nilai b , semakin besar kemampuan yang dibutuhkan bagi responden untuk memiliki kemungkinan 50% menjawab benar, artinya aitem tersebut semakin sulit (Hambleton, Swaminathan, Rogers, 1991). Aitem yang sulit terletak di sebelah kanan atau ujung skala kemampuan yang lebih tinggi, aitem mudah terletak di kiri atau ujung bawah dari skala kemampuan (Hambleton, Swaminathan, Rogers, 1991).

Indeks kesukaran aitem dalam pengukuran *Rasch Model* dapat dilihat melalui nilai *logit* pada tabel *item measure*, dan sudahurut dari nilai *logit* paling tinggi hingga paling rendah (Sumintono & Widhiarso, 2014, 2015). Nilai *logit* yang tinggi menunjukkan kesukaran aitem juga tinggi (Sumintono & Widhiarso, 2015). Indeks kesukaran aitem disimbolkan dengan b . Nilai b bergerak dari -2 hingga +2 (Hambleton dan Swaminathan, 1985).

Pada *item measure* terdapat informasi nilai standar deviasi, yang apabila dikombinasikan dengan rata-rata mean *logit*, maka tingkat kesukaran aitem bisa

B. Subtes *Satz Ergaenzung* (SE)

1. Pengertian Subtes SE

Satz Ergaenzung (SE) merupakan subtes pertama dari sembilan subtes pada *Intelligenz Struktur Test* (IST). Berdasarkan kamus *Universal Langenscheidt* (Adiwimarta, Darmojuwono, Hastrick, 2011), kata *satz* berarti kalimat dan *ergaenzen* berarti melengkapi, maka jika diterjemahkan *satz ergaenzung* artinya melengkapi kalimat. SE mengukur kemampuan *reasoning*, yang berkaitan dengan pembentukan keputusan, *common sense* (memanfaatkan pengalaman masa lalu), dan berpikir konkrit praktis terhadap apa yang dialami sehari-hari (Wiratna, 1993). Aitem-aitemnya akan menuntut subjek untuk melakukan penilaian berdasarkan pengalaman konkrit dan informasi faktual yang dimilikinya.

Penalaran (*reasoning*) adalah kegiatan berpikir seturut asas kelurusan berpikir atau sesuai dengan hukum logika (Sobur, 2003). Bernalar tidak sama dengan berpikir. Penalaran memang merupakan aspek penting dalam berpikir (Sobur, 2003). Menurut Sudarminta, bernalar adalah kegiatan pikiran untuk menarik kesimpulan dari premis-premis yang sebelumnya sudah diketahui (Sobur, 2003: 209). Penalaran dapat dibagi menjadi dua kategori, yaitu penalaran deduktif dan penalaran induktif (Ling & Catling, 2012; Sobur, 2003, Solso, dkk., 2008). Penalaran deduktif merupakan proses pengambilan keputusan dari umum ke khusus (Ling & Catling, 2012). Penalaran induktif merupakan perpindahan fakta-fakta spesifik ke suatu kesimpulan (Ling & Catling, 2012).

Kemampuan *reasoning* yang diukur oleh SE tergolong pada penalaran induktif. Pada penalaran ini, suatu kesimpulan biasanya dinyatakan secara implisit



atau eksplisit dalam konteks pernyataan kemungkinan (Solso, dkk., 2008). Keputusan yang diambil melalui penalaran induktif, berdasar pada pengalaman masa lalu dan kesimpulannya berdasarkan yang dirasa sebagai pilihan terbaik dari sejumlah alternatif (Solso, dkk., 2008). Hal ini juga sejalan dengan berbagai definisi penalaran yang berfokus pada keputusan atau kesimpulan yang dibuat. Tversky menyatakan bahwa dalam mengambil keputusan kita memilih alternatif dengan cara mengeliminasi pilihan yang kurang menarik secara bertahap (Solso, dkk., 2008).

Pengambilan keputusan melibatkan proses penarikan kesimpulan, dapat dilakukan dengan bermacam cara, yaitu:

- a. Penarikan kesimpulan atas dasar analogi (Walgito, 2010)

Kesimpulan yang ditarik atas dasar adanya kesamaan dari suatu keadaan atau peristiwa dengan keadaan atau peristiwa yang lain.

- b. Penarikan kesimpulan atas dasar cara induktif (Sobur, 2003; Solso, dkk., 2008; Walgito, 2010)

Menarik kesimpulan umum dari berbagai kejadian yang ada di sekitarnya. Keputusan yang diambil berdasarkan pengalaman masa lalu dan kesimpulannya berdasarkan pilihan yang dirasa paling baik dari sejumlah alternatif.

- c. Penarikan kesimpulan atas dasar cara deduktif (Sobur, 2003; Walgito, 2010)

Kesimpulan yang ditarik atas dasar hal yang umum ke hal yang bersifat khusus. Jalannya pengambilan keputusan yaitu ada ketentuan, kemudian dikenakan pada peristiwa.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2. Karakteristik Subtes SE

Subtes SE disusun pada bagian pertama dalam buku tes IST, oleh karena itu disebut sebagai subtes pertama. Subtes SE terdiri dari 20 soal berupa kalimat-kalimat, dengan lima pilihan jawaban, sehingga disebut subtes melengkapi kalimat (Wiratna, 1993). Waktu pengerjaan untuk subtes SE adalah 6 menit. Skoring subtes SE berupa dikotomi, yaitu skor 0 untuk jawaban salah dan skor 1 untuk jawaban benar (Wiratna, 1993). Subtes ini mengharuskan subjek untuk memilih salah satu kata yang tepat untuk mengisi satu kata yang hilang, sehingga susunan kalimat dalam soal menjadi sempurna.

C. Kerangka Berpikir

Setiap tes psikologi haruslah memiliki kualitas yang baik. Kualitas yang dimaksud ini dilihat dari karakteristik pengukuran tes tersebut, yang paling utama adalah validitas dan reliabilitas tes (Azwar, 2016). Validitas menunjukkan bahwa tes mampu memperoleh hasil pengukuran sesuai dengan tujuannya, sedangkan reliabilitas menunjukkan konsistensi hasil yang diberikan oleh tes bila digunakan secara berulang. Alat tes yang valid dan reliabel inilah yang dapat digunakan untuk memperoleh informasi akurat.

Subtes *satzer gaenzung* (SE) pada IST merupakan instrumen untuk mengukur kemampuan *reasoning*, sebagai bagian dari komponen inteligensi, yang masih sering digunakan. Penggunaan SE ini biasanya disajikan satu paket dengan IST. Tes ini digunakan untuk keperluan seleksi dan penempatan di dunia kerja dan juga



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

pendidikan, seperti yang digunakan oleh Biro Layanan Psikologi Atma Jaya, P3M Fakultas Psikologi Sumatera Utara, dan BKPH Fakultas Psikologi UIN Suska Riau.

Sebagai alat tes yang banyak digunakan di Indonesia, SE haruslah memiliki kualitas baik, agar memberikan hasil yang akurat. Tes ini digunakan sejak 64 tahun lalu dan sudah mengalami revisi berkali-kali dalam perjalanannya hingga saat ini, walaupun versi yang digunakan di Indonesia bukanlah versi terbarunya. Di samping itu, berdasarkan temuan dari beberapa penelitian dalam kurun waktu kurang dari 10 tahun belakangan, SE justru memiliki kualitas yang kurang baik untuk terus eksis digunakan. Penelitian dari Widianti (2008) dan Elvira (2011) menunjukkan bahwa SE tidak mengukur konstruk yang spesifik, serta konsistensi dari tes ini belum baik. Hasil penelitian Agung dan Fitri (2016) menyatakan bahwa SE merupakan subtes yang memiliki reliabilitas paling rendah dalam kategori kurang baik, dibandingkan dengan delapan subtes IST lainnya.

Pendekatan yang digunakan untuk mengestimasi karakteristik pengukuran suatu instrumen juga mempengaruhi hasilnya, baik itu pendekatan teori tes klasik maupun *item response theory* (IRT). Berdasarkan hasil penelitian Elvira (2011) dan Rahmawati (2014) menggunakan data yang sama dari P3M Fakultas Psikologi Universitas Sumatera Utara, bahwasannya estimasi properti psikometri dengan teori tes klasik hanya menghasilkan satu aitem yang buruk, sedangkan estimasi menggunakan IRT, menghasilkan sembilan aitem yang buruk. Melalui estimasi menggunakan IRT, ada lebih banyak aitem yang direduksi dari SE. Hal ini berarti bahwa IRT lebih teliti dalam mengestimasi karakteristik psikometri, sehingga hasilnya akan lebih tepat (Sumintono & Widhiarso, 2014).



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau
 Stee Islam University Sultan Syarif Kasim Riau

Pada studi awal yang penulis lakukan dengan mengestimasi karakteristik psikometri dari 293 data hasil tes SE mahasiswa tahun pertama UIN Suska Riau, menggunakan pendekatan *Rasch Model*, terlihat bahwa reliabilitas SE tergolong sangat buruk. Melalui uji parameter butir soal, penulis menemukan adanya satu aitem *misfit*, lima aitem dengan indeks kesukaran yang tidak baik, serta satu aitem bias. Pengujian awal ini dilakukan untuk membuktikan bahwa SE yang digunakan saat ini memiliki kualitas yang kurang memuaskan secara properti psikometri.

Estimasi terhadap karakteristik pengukuran sebuah tes menginformasikan apakah alat tes tersebut masih dapat digunakan atau tidak (Azwar, 2011). Karakteristik psikometri tes dan aitem tes yang dihasilkan, memberikan bantuan kepada kita untuk melakukan perbaikan pada aitem dan tes yang berkualitas buruk, sehingga dapat meningkatkan kualitas aitem dan tes tersebut. Berdasarkan penelitian terdahulu dan pengujian awal yang telah penulis lakukan, maka sudah saatnya kini melakukan revisi terhadap aitem-aitem SE. Agung dan Fitri (2016), Elvira (2011) dan Rahmawati (2014) juga menyarankan untuk melakukan revisi atau modifikasi pada SE guna memperbaiki kualitas tes ini. Oleh karena itu, penelitian ini akan melakukan revisi pada aitem-aitem SE berdasarkan hasil studi awal dan melakukan pengujian karakteristik psikometri menggunakan pendekatan *Rasch Model* pada SE revisi.

D. Hipotesis

Hipotesis pada penelitian ini adalah karakteristik psikometri subtes *satz ergaenzung* (SE) revisi sudah memenuhi kriteria unidimensionalitas, reliabilitas, *item fit*, indeks kesukaran aitem, dan DIF.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.