

Aplikasi dan Praktik Konstruksi Skala Psikologi sebagai Instrumen Pengukuran dalam Penelitian Pendidikan

Alfrits Roul Sinadia

Fakultas Keguruan dan Ilmu pendidikan, Universitas Klabat, Manado,
alfritssinadia@unklab.ac.id

Abstrak

Latar belakang diadakannya penelitian ini ialah adanya miskonsepsi definisi dan fungsi dari angket dan kuesioner yang biasa digunakan untuk mengukur atribut psikologi dalam penelitian pendidikan. Penelitian ini merupakan sebuah studi literatur yang mereview dan mensintesiskan berbagai tulisan atau teori-teori pengembangan skala psikologi dalam praktik konstruksi instrumen yang digunakan dalam penelitian pendidikan. Hasil penelitian menunjukkan bahwa dalam pengembangan skala psikologi, peneliti perlu memastikan bahwa skala yang dikembangkan dapat dibuktikan validitas dan diestimasi reliabilitasnya. Secara spesifik, peneliti wajib menyusun kisi-kisi (blueprint) skala, merumuskan komponen, indikator, dan subindikator skala tersebut. Selain itu, kajian literatur menunjukkan bahwa pembuktian validitas skala psikologi dapat dilakukan dengan analisis Aiken's V dan estimasi reliabilitas yang sesuai ialah analisis coefficient alpha.

Kata kunci—pengukuran atribut psikologi, penelitian pendidikan, skala psikologi, instrumen pengukuran

Abstract

The background of this study is the misconception of the definition and function of psychology scales and questionnaires that are commonly used to measure psychological attributes in educational research. This study is a literature study that reviews and synthesizes various writings or theories of psychological scale development in the practice of constructing instruments used in educational research. The results of the study show that in the development of psychological scales, researchers need to ensure that the scale developed can be proven valid and estimated to be reliable. Specifically, researchers are required to prepare a scale grid (blueprint), formulate components, indicators, and subindicators of the scale. In addition, literature review shows that proving the validity of the psychological scale can be done by Aiken's V analysis and the appropriate reliability estimation is coefficient alpha analysis.

Keywords—measurement of psychological attributes, educational research, psychological scale, measurement instruments

PENDAHULUAN

Pengukuran atribut psikologi merupakan salah satu dari berbagai objek yang diteliti dalam bidang pendidikan. Sebagai contoh, objek penelitian dalam bidang pendidikan dapat menyasar atribut motivasi belajar, minat belajar, sikap dalam belajar, kecerdasan dalam belajar, dan atribut lainnya yang sesungguhnya merupakan aplikasi dari berbagai teori psikologi dalam dunia pendidikan. Secara garis besar, atribut psikologi dapat dibagi menjadi dua kategori, yaitu atribut kemampuan (kognitif) dan atribut bukan kemampuan (Azwar, 2017b). Atribut kemampuan atau kognitif dapat dibagi lagi menjadi kemampuan potensial seperti inteligensi dan bakat dan kemampuan yang bersifat aktual seperti prestasi. Di sisi lain, atribut bukan kemampuan inilah yang merujuk kepada atribut psikologi seperti minat, sikap, kecemasan, dan lain-lain. Kedua kategori atribut tersebut umumnya merupakan data yang sering dikumpulkan untuk dianalisis dalam penelitian-penelitian psikologi di bidang pendidikan.

Terkait dengan jenis data yang dikumpulkan dalam penelitian-penelitian psikologi di bidang pendidikan, (Azwar, 2017a) menggolongkan jenis data ini menjadi dua jenis, yaitu data faktual dan data non faktual. Data faktual merupakan data yang diketahui oleh subjek yang diteliti seperti usia, frekuensi kehadiran, tingkat pendidikan, dan lain-lain. Di sisi lain, ada data yang bersifat non faktual karena subjek penelitian itu sendiri tidak mengetahui fakta yang sedang diteliti tersebut. Data ini disebut oleh (Azwar, 2017a) sebagai data variabel dari subjek penelitian yang hanya dapat diungkap melalui pengukuran. Sebagai contoh, data variabel ini dapat berupa kecemasan, motivasi belajar, kestabilan emosi, sikap, dan lain-lain. Dalam praktik penelitian psikologi pendidikan di lapangan, kedua jenis data ini umum dikumpulkan dalam berbagai penelitian pendidikan

Dalam penelitian-penelitian pendidikan yang menyasar atribut psikologi, salah satu alat ukur yang umum digunakan untuk pengumpulan data ialah kuesioner. Umumnya, sebuah kuesioner dikenal sebagai alat yang digunakan untuk mengumpulkan informasi mengenai pemikiran, perasaan, sikap, kepercayaan, nilai-nilai, persepsi, kepribadian, dan maksud-maksud perilaku lainnya dari partisipan dalam sebuah penelitian (Johnson & Christensen, 2014). Namun, istilah kuesioner secara konseptual dari sudut pandang penelitian psikologi sesungguhnya digunakan untuk mengumpulkan jenis informasi yang berbeda dari yang dinyatakan oleh Johnson dan Christensen. (Azwar, 2017b) secara konseptual membedakan antara istilah instrumen yang digunakan dalam penelitian psikologi untuk pengumpulan data faktual dan non faktual. Ia menjelaskan bahwa sebenarnya angket (kuesioner) digunakan sebagai alat untuk mengumpulkan data faktual, dan sebaliknya tes serta skala psikologi digunakan sebagai alat untuk mengumpulkan data non faktual. Secara spesifik, Azwar menyatakan bahwa dalam praktik penelitian, instrumen yang diistilahkan sebagai tes merujuk kepada alat ukur kemampuan kognitif, sedangkan istilah skala lebih banyak digunakan untuk menyebut alat ukur atribut non kognitif atau yang juga disebut skala psikologi.

HASIL DAN DISKUSI

Merujuk pada istilah konseptual instrumen yang digunakan oleh (Azwar, 2017b), teridentifikasi sebuah miskonsepsi dalam praktik pengumpulan data penelitian psikologi di bidang pendidikan. Miskonsepsi ini teramati berdasarkan hasil pengamatan peneliti di lapangan selama bertahun-tahun sebagai tenaga pengajar di tingkat universitas, membimbing penulisan skripsi mahasiswa, dan melakukan penelitian sebagai dosen di Fakultas Keguruan dan Ilmu Pendidikan. Miskonsepsi ini nampaknya telah tumbuh dan berakar dalam selama bertahun-tahun dalam praktik pengerjaan penelitian mahasiswa.

Dalam praktik penelitian pemula seperti yang umum dilakukan oleh mahasiswa tingkat sarjana, angket (kuesioner) dianggap merupakan alat untuk mengumpulkan semua jenis data baik data faktual maupun data nonfaktual. Pemahaman para mahasiswa tingkat sarjana nampaknya masih mengaburkan konsep antara istilah angket (kuesioner) dan skala

psikologi. Mungkin ini disebabkan oleh kurangnya pemahaman mereka terhadap perbedaan konsep kedua bentuk instrumen tersebut, atau kurangnya literatur terkait yang dapat diakses oleh para mahasiswa tersebut, yang menjelaskan mengenai perbedaan ini.

Selain itu, hasil pengamatan peneliti pada proses pengerjaan skripsi mahasiswa di lapangan menunjukkan bahwa mahasiswa tingkat sarjana seringkali memiliki pemahaman yang kurang mengenai proses konstruksi instrument psikologi dalam penelitian pendidikan. Proses konstruksi skala psikologi dalam penelitian mereka terkesan asal-asalan dibuat dan belum dilakukan sesuai prinsip-prinsip pengembangan sebuah instrumen yang berkualitas. Hal ini mungkin juga disebabkan oleh kurangnya akses kepada sumber-sumber literatur atau buku-buku yang relevan dengan penelitian alat ukur psikologi.

Oleh sebab-sebab yang telah dipaparkan sebelumnya, studi literatur ini bertujuan untuk mengungkap perbedaan konsep antara istilah angket (kuesioner) dan skala psikologi. Kedua, studi ini juga bertujuan untuk mereview proses pengembangan sebuah skala psikologi yang sesuai dengan konsep-konsep pengukuran psikologi (psikometri). Selain itu, hasil studi literatur ini dapat menjadi panduan ringkas bagi para peneliti pemula dalam pengembangan alat ukur atribut psikologi. Dengan demikian, konstruksi skala psikologi yang umum dilakukan oleh para mahasiswa tingkat sarjana dapat didasarkan pada teknik-teknik penelitian alat ukur psikologi yang telah umum diterima dan dipraktikkan dalam bidang ilmu pengukuran psikologi atau psikometri.

Karakteristik Skala Psikologi

Menurut (Azwar, 2017b), skala psikologi memiliki beberapa karakteristik. Pertama, butir yang dituliskan pada skala psikologi merupakan pertanyaan atau pernyataan yang bertujuan untuk mengukur atribut perilaku dari subjek (responden) secara tidak langsung. Umumnya, subjek atau responden akan menjawab butir ini dengan memilih pilihan jawaban yang lebih sesuai menggambarkan tentang diri dan perasaannya. Kedua, skala psikologi selalu berisi banyak butir. Hal ini disebabkan oleh karena atribut dalam diri subjek hanya diungkap secara tidak langsung berdasarkan indikator-indikator perilaku yang dikemas dalam bentuk butir-butir pertanyaan atau pernyataan. Untuk menarik kesimpulan, peneliti tidak hanya menganalisis respons subjek atau responden terhadap satu butir saja, namun perlu menganalisis respons terhadap keseluruhan butir yang terdapat pada skala psikologi tersebut. Ketiga, respons atau jawaban dari subjek tidak dapat dikategorikan sebagai respons benar atau salah. Semua jawaban yang diberikan subjek dapat diterima karena itu hanyalah skor yang mewakili atribut yang diukur, misalnya 5, 4, 3, 2, atau 1. Oleh sebab karakteristik ini, (Azwar, 2017b) kemudian menyatakan bahwa konsistensi hasil pengukuran dari atribut yang diukur tergantung kepada kejujuran subjek (responden) yang menjawab butir-butir pertanyaan tersebut.

Prinsip-Prinsip Konstruksi Skala Psikologi

Skala psikologi perlu dikembangkan secara seksama dan hati-hati. Guna menghasilkan skala psikologi yang berkualitas tinggi, skala tersebut perlu dikembangkan melalui langkah-langkah yang sesuai dengan teori-teori pengukuran psikologi. Secara konseptual, skala psikologi yang dikonstruksi perlu memenuhi prinsip validitas dan reliabilitas instrumen (Mardapi, 2008; Reynolds dkk., 2009; Azwar, 2014) serta melalui proses analisis butir (Reynolds et al., 2009). Para peneliti yang memilih untuk mengembangkan skala psikologi mereka sendiri, sepatutnya memastikan bahwa alat ukur yang mereka kembangkan memenuhi prinsip-prinsip validitas dan reliabilitas alat ukur. Selain kedua prinsip ini, hal lain yang perlu diperhatikan oleh para peneliti skala psikologi ialah kualitas butir yang terdapat pada alat ukur tersebut. Kualitas alat ukur secara keseluruhan tergantung pada kualitas setiap butir pada instrumen tersebut (Reynolds et al., 2009). Validitas atau ketepatan penafsiran hasil pengukuran dan reliabilitas atau konsistensi

hasil pengukuran tergantung pada kualitas butir di instrumen tersebut. Untuk itu, butir-butir yang dikembangkan secara seksama dan hati-hati akan menghasilkan skala psikologi yang berkualitas tinggi.

Validitas Skala Psikologi

Validitas hasil pengukuran dari sebuah skala psikologi merupakan keharusan yang wajib dipenuhi oleh para peneliti instrumen ini. Secara umum, validitas didefinisikan sebagai pertimbangan atau perkiraan seberapa baiknya sebuah alat ukur mengukur apa yang dimaksudkan akan diukurnya (Swerdlik, 2009). Serupa dengan itu, validitas diartikan sebagai akurasi dan kecermatan sebuah instrumen dalam menjalankan fungsinya sebagai alat ukur (Azwar, 2017b). Namun, bila ditinjau dari karakteristik tafsiran skor, maka validitas merujuk kepada kesesuaian atau akurasi tafsiran dari skor yang dihasilkan oleh alat ukur tersebut (Reynolds et al., 2009). Hal ini sejalan dengan arti validitas yang didefinisikan oleh AERA dkk. (2014) bahwa validitas merujuk kepada tingkat di mana bukti dan teori mendukung tafsiran skor yang dihasilkan oleh alat ukur untuk tujuan pengukuran yang telah ditetapkan. Ini berarti bahwa sebuah alat ukur yang tepat akan menghasilkan karakteristik tafsiran skor yang tepat.

Validitas Isi. Terdapat beberapa cara yang dapat ditempuh untuk membuktikan validitas dari skala psikologi. Mardapi (2008) mengusulkan salah satu cara pembuktian, yaitu pembuktian berdasarkan isi alat ukur atau validitas isi (*content validity*). Ia kemudian menjelaskan bahwa bukti berdasarkan isi alat ukur atau validitas isi terdiri dari dua tipe validitas, yaitu validitas tampang (*face validity*) dan validitas logik (*logical validity*). Validitas tampang umumnya dilakukan dengan memeriksa kesesuaian butir-butir yang ditulis dengan aspek psikologi yang diukurnya. Di sisi lain, validitas logik merupakan batasan kawasan atribut yang diukur atau sejauh mana butir-butir yang terdapat dalam alat ukur tersebut telah menjangkau keseluruhan batasan kawasan atribut yang diukur.

Dalam praktik penelitian skala psikologi di lapangan, pembuktian validitas isi (*content validity*) dapat dilakukan dengan meminta pendapat para ahli mengenai isi alat ukur tersebut. Sebelum isi alat ukur tersebut diajukan kepada para ahli, peneliti perlu melakukan beberapa langkah pengembangan. Langkah pertama, peneliti perlu membuat kisi-kisi (*blueprint*) seperti yang disarankan oleh Azwar (2017b) (lihat Tabel 1). Dalam pembuatan kisi-kisi ini, peneliti perlu merumuskan aspek berperilaku dari atribut yang diukur, lalu merumuskan indikator-indikator dari setiap aspek berperilaku. Proses perumusan aspek dan indikatornya ini didasarkan pada teori-teori psikologi yang mendasari skala tersebut. Setiap indikator yang ada diberi bobot agar peneliti dapat mengetahui dengan jelas indikator mana yang perlu mendapatkan penekanan yang lebih

Tabel 1. Kisi-Kisi Skala Regulasi Diri

Aspek	Indikator	Bobot (%)
A. Mampu Mengendalikan Emosi	1. Tidak reaktif terhadap stimulus negatif (tidak dikehendaki)	20
	2. Dapat menahan atau menunda ekspresi emosi	10
	3. Lebih mengutamakan logika daripada perasaan dalam menghadapi masalah	20
B. Memiliki Inisiatif dalam Bekerja	4. Dapat mencari/menggunakan cara berbeda untuk menyelesaikan pekerjaan	20
	5. Berani bertanya/berpendapat demi hasil yang lebih baik	15
	6. Melaksanakan tugas pada waktunya tanpa menunggu perintah	15
Total		100%

besar atau sebaliknya. Sebagai contoh, jika peneliti ingin membuat skala psikologi yang terdiri dari 30 butir, maka untuk indikator 1 diperlukan sebanyak enam butir pertanyaan/pernyataan untuk mengukurnya. Jumlah ini diperoleh melalui perhitungan berikut: jumlah persentase dikali total butir, dibagi seratus ($20 \times 30 / 100 = 6$). Untuk indikator 2 dalam contoh di atas, diperlukan setidaknya tiga butir pertanyaan/pernyataan untuk mengukurnya yang diperoleh dari perhitungan $10 \times 30 / 100 = 3$. Jumlah butir yang diperlukan dalam sebuah skala psikologi dapat disesuaikan dengan kebutuhan pengukuran.

Selain contoh yang diberikan oleh Azwar (2017b), peneliti juga dapat mencoba pembuatan kisi-kisi skala seperti yang dicontohkan oleh Mardapi (2008) (lihat Tabel 2).

Tabel 2. Kisi-Kisi Alat Ukur Sikap

Definisi Konseptual	: Sikap adalah			
Definisi Operasional	:			
No	Indikator	Jumlah Butir	Pertanyaan/Pernyataan	Skala
1				
2				
3				
4				

Dalam contoh di atas, peneliti dapat menuliskan sendiri definisi konseptual dan operasional dari atribut yang akan diukur. Setelah itu, barulah dituliskan indikator-indikator yang muncul dari atribut tersebut.

Dalam penyusunan indikator-indikator atribut psikologi, peneliti dapat mengkonstruksinya dengan mengkaji teori-teori psikologi yang relevan dengan atribut tersebut (Retnawati, 2016). Untuk memperjelas setiap indikator, Retnawati menjelaskan bahwa peneliti dapat mengembangkan subindikator dari setiap indikator yang ada, lalu menyusun butir instrumen untuk mengukur setiap subindikator tersebut. Penyusunan kisi-kisi seperti ini dapat dilihat seperti dalam contoh yang dikembangkan oleh Retnawati (2016) berdasarkan Zimmerman (2000) (lihat Tabel 3).

Tabel 3. Komponen dan Indikator Self-Regulated Learning

Komponen	Indikator	Sub Indikator	No. Butir
Pemikiran	Analisis Tugas	Pengaturan Tujuan	1
		Perencanaan Strategis	2
	Keyakinan Diri	Kemampuan Diri	3
		Orientasi Tugas	4
Kontrol Kinerja	Pengendalian Diri	Instruksi Diri	5
		Usaha untuk Fokus Belajar	6
		Strategi Penyelesaian Tugas	7
	Pengamatan yang Cukup	Pemantauan Metakognitif	8
		Catatan Diri	9
		Ekperimentasi Diri	10
	Pertimbangan Diri	Evaluasi Diri	11
		Atribusi Kausal	12
	Reaksi Diri	Kepuasan Diri (Hadiah)	13
		Kepuasan Diri (Hukuman)	14
		Adaptif/Defensif	15

Langkah kedua, dengan menggunakan komponen, indikator, dan subindikator yang telah dirumuskan dalam kisi-kisi, peneliti dapat mulai menuliskan butir-butir skala psikologi yang direncanakan. Sebagai contoh, berdasarkan kisi-kisi pada Tabel 3, Retnawati (2016) kemudian menuliskan butir-butir untuk mengukur *self-regulated learning*. Setiap subindikator diukur oleh satu butir pernyataan sehingga total terdapat 15 butir pernyataan

untuk skala *self-regulated learning* (lihat Tabel 4). Pilihan jawaban yang disediakan untuk responden dituliskan dalam bentuk skala Likert dengan empat pilihan, yaitu TP (tidak pernah), J (jarang), S (sering), dan SL (selalu).

Tabel 4. Butir untuk Mengukur Self-Regulated Learning dengan Likert

No	Pernyataan	TP	J	S	SL
1	Saya merumuskan tujuan-tujuan kuliah/belajar saya sebelum kegiatan dimulai.				
2	Saya merencanakan strategi untuk mencapai tujuan kuliah/belajar saya.				
3	Saya mempercayai kemampuan diri saya untuk berhasil dalam kuliah/belajar.				
4	Saya menitikberatkan usaha mencapai tujuan kuliah/belajar saya dibandingkan dengan kegiatan lain.				
5	Saya membuat jadwal untuk diri sendiri terkait dengan pencapaian tujuan kuliah/belajar saya.				
6	Saya mengupayakan diri untuk fokus belajar.				
7	Saya menyusun strategi paling tepat untuk penyelesaian tugas kuliah/belajar.				
8	Saya membuat peta kegiatan/aktivitas yang telah saya lakukan.				
9	Saya membuat catatan apa yang telah saya lakukan baik yang berhasil maupun yang belum.				
10	Jika ada hal yang membuat saya gagal, saya akan berusaha lagi dengan strategi lain.				
11	Setelah selesai melakukan kegiatan dan melihat hasilnya (misal akhir semester) saya melakukan evaluasi.				
12	Saya mencermati penyebab keberhasilan atau kegagalan usaha saya.				
13	Setelah mencapai hal sesuai target kuliah/belajar, saya memberi hadiah untuk diri sendiri.				
14	Saya menghukum diri sendiri jika ada hal dari diri sendiri yang menyebabkab saya gagal mencapai target kuliah/belajar.				
15	Jika suatu strategi kuliah/belajar yang saya gunakan berhasil, saya akan menggunakannya lagi.				

Langkah ketiga, setelah penulisan butir-butir skala psikologi selesai dilakukan, peneliti dapat mengajukan draft skala psikologi yang telah dibuat kepada beberapa ahli yang relevan dengan atribut yang akan diukur. Dalam hal ini, menurut Retnawati (2016), peneliti dapat meminta sekurang-kurangnya tiga orang ahli untuk memeriksa instrumen yang telah dikonstruksi oleh peneliti. Para ahli kemudian dapat menilai kesesuaian atau relevansi setiap butir dengan indikatornya dengan menggunakan contoh penilaian yang didesain oleh Retnawati (lihat Tabel 5). Setiap penilai ahli atau rater akan memberikan penilaian dalam bentuk skor berikut: 5 = sangat relevan, 4 = relevan, 3 = cukup relevan, 2 = kurang relevan, 1 = tidak relevan. Selain memberikan skor, penilai juga dapat menuliskan saran-saran perbaikan pada tabel penilaian instrumen.

Tabel 5. Format Penilaian Ahli untuk Mengetahui Kesesuaian Butir dengan Indikator

No Butir	Skor Relevansi Butir dengan Indikator					Keterangan
	1	2	3	4	5	
	Tidak Relevan	Kuran Relevan	Cukup Relevan	Relevan	Sangat Relevan	
1						
2						
3						

Skor atau rating yang diberikan oleh para penilai kemudian dianalisis untuk menentukan validitas isi dari butir-butir yang telah dikonstruksi. Untuk membuktikan validitas isi, peneliti dapat menggunakan analisis Aiken's V (Aiken, 1985) seperti yang disarankan oleh Azwar (2014). Nilai V dapat dihitung menggunakan rumus berikut:

$$V = \frac{\sum s}{[n(c - 1)]}$$

s = r - lo

lo = Angka penilaian validitas terendah (dalam hal ini = 1)

c = Angka penilaian validitas yang tertinggi (dalam hal ini = 5)

r = Angka yang diberikan seorang penilai

Nilai V dari setiap butir berkisar antara 0,00 sampai 1,00 (Azwar, 2014) dan dapat ditafsirkan menggunakan skala berikut: < 0,4 = validitas rendah; 0,4-0,8 = validitas sedang; > 0,8 = validitas tinggi (Retnawati, 2016). Jika sebuah butir memiliki validitas isi rendah, maka peneliti dapat memutuskan untuk mengeluarkan butir tersebut dari instrumen atau melakukan perbaikan pada butir tersebut untuk dinilai kembali.

Validitas Konstruk. Cara kedua untuk membuktikan validitas dari skala psikologi ialah dengan menggunakan teknik validitas konstruk (*construct validity*). Validitas konstruk merupakan pertimbangan kesesuaian dari kesimpulan yang ditarik dari hasil pengukuran posisi individual pada sebuah variabel yang disebut konstruk (Swerdlik, 2009) atau variabel laten, yaitu variabel yang tidak dapat diukur secara langsung (Field, 2009). Oleh Mardapi (2008), validitas ini disebut sebagai bukti berdasarkan pola respons. Validitas konstruk menurut Mardapi mengacu pada sejauh mana suatu alat ukur mengukur konsep dari suatu teori yang menjadi dasar konstruksi dari sebuah skala psikologi. Dengan melakukan teknik validitas konstruk, peneliti dapat membuktikan kesesuaian butir-butir yang dikonstruksi dengan komponen teori atau indikator yang telah dirumuskan di tahapan validitas isi.

Teknik analisis yang dapat dilakukan oleh peneliti untuk menganalisis validitas konstruk dari skala psikologi ialah melalui analisis faktor. Istilah faktor itu sendiri berasal dari penyebutan lain terhadap variabel laten (Field, 2009) atau yang disebut atribut psikologi (Azwar, 2017b). Analisis faktor itu sendiri terbagi menjadi dua, yaitu analisis faktor eksploratori (*exploratory factor analysis*) dan analisis faktor konfirmatori (*confirmatory factor analysis*) (Retnawati, 2016). Analisis faktor eksploratori cenderung digunakan di tahap-tahap awal penelitian untuk menemukan informasi tentang hubungan antara satu set variabel (Pallant, 2007). Pallant kemudian menjelaskan bahwa teknik yang kedua, analisis faktor konfirmatori, adalah teknik yang lebih kompleks yang biasanya digunakan di tahap selanjutnya dalam sebuah penelitian untuk memastikan hipotesis atau teori yang terkait dengan struktur yang mendasari set variabel tersebut. Kedua jenis analisis ini merupakan analisis validitas konstruk yang umum digunakan oleh para peneliti di bidang psikologi dan pendidikan.

Untuk kemudahan dalam melakukan analisis faktor, peneliti dapat menggunakan bantuan perangkat lunak komputer. Saat ini sudah terdapat banyak program komputer yang dapat digunakan untuk melakukan berbagai analisis ini. Sebagai contoh, analisis faktor eksploratori dapat dilakukan menggunakan SPSS (Pallant, 2007; Field, 2009), atau menggunakan SAS, MINITAB, R, MPLUS, dan lain-lain (Retnawati, 2016). Di sisi lain, analisis faktor konfirmatori dapat dilakukan dengan bantuan perangkat lunak seperti Lisrel, AMOS, MPLUS, PLS, atau aplikasi lain (Retnawati, 2016). Dengan demikian, analisis validitas konstruk melalui analisis faktor tidak perlu dilakukan secara manual karena para peneliti dimudahkan dengan penggunaan perangkat lunak komputer.

Reliabilitas Skala Psikologi

Di samping validitas, skala psikologi juga perlu memiliki reliabilitas yang baik. Dalam bahasa pengukuran psikologi, reliabilitas merujuk kepada konsistensi dalam pengukuran (Swerdlik, 2009), atau yang berarti seberapa tinggi kecermatan pengukuran (Azwar, 2017b). Azwar berargumen bahwa karakteristik dari alat ukur psikologi yang berkualitas baik ialah alat ukur tersebut mampu menghasilkan skor yang cermat dengan kesalahan (error) pengukuran yang kecil. Semakin kecil error yang terdapat dalam hasil pengukuran sebuah skala psikologi, konsistensi hasil pengukuran akan semakin tinggi. Sebaliknya, semakin besar error yang dihasilkan, semakin rendah konsistensi dalam pengukuran menggunakan skala tersebut.

Tingkat reliabilitas atau kestabilan suatu skala psikologi dinyatakan dalam bentuk indeks reliabilitas atau koefisien reliabilitas. Koefisien reliabilitas secara teoretik berkisar antara 0,00 hingga 1,00 (Azwar, 2014). Dalam praktik pengukuran, koefisien reliabilitas yang baik ialah yang nilainya mendekati 1,00. Semakin mendekati 1,00, koefisien reliabilitas dari sebuah skala psikologi akan semakin baik karena kesalahan pengukuran yang dihasilkan akan semakin kecil. Menurut Reynolds dkk. (2009), jika sebuah skala psikologi memiliki koefisien reliabilitas sebesar 0,90, ini berarti bahwa 90% variansi hasil pengukuran merupakan variansi yang tepat, dan sisanya 10% merupakan variansi hasil pengukuran yang mengandung error. Sebagai contoh lain, jika sebuah skala psikologi memiliki koefisien reliabilitas sebesar 0,60, variansi hasil pengukuran yang mengandung kesalahan dari skala tersebut ialah sebesar 40%. Para peneliti sebaiknya tidak menggunakan skala dengan tingkat error sebesar itu. Untuk alasan ini, (Reynolds et al., 2009) menyarankan agar koefisien reliabilitas dari sebuah skala psikologi tidak kurang dari atau minimal 0,70. Dengan koefisien reliabilitas sebesar ini, maka estimasi variansi hasil pengukuran yang tepat dari skala psikologi yang digunakan ialah sebesar 70% dan error yang dihasilkan hanya sebesar 30%. Dapat disimpulkan bahwa semakin kecil variansi error yang dihasilkan oleh sebuah skala psikologi, semakin reliabel (konsisten) skala tersebut untuk mengukur atribut psikologi yang akan diukur.

Terdapat beberapa kategori analisis reliabilitas yang dapat digunakan oleh peneliti untuk menghitung indeks atau koefisien reliabilitas sebuah instrumen atau alat ukur. (Reynolds et al., 2009) dan (Swerdlik, 2009) mendaftarkan beberapa tipe reliabilitas, yaitu *test-retest*, *alternate-forms*, *internal consistency*, dan *inter-scorer* atau *inter-rater*. Manakah dari tipe reliabilitas ini yang sesuai untuk digunakan pada sebuah skala psikologi? Pertanyaan ini seringkali ditanyakan oleh para peneliti pemula saat memilih tipe reliabilitas yang sesuai untuk digunakan. Umumnya praktik administrasi di lapangan atau pemberian skala psikologi kepada responden untuk diisi ialah pemberian sekali saja. Dengan kata lain, responden hanya mengisi skala yang diberikan kepada mereka sekali saja dan peneliti kemudian mengambil data tersebut untuk dianalisis. Untuk praktik seperti ini, maka (Swerdlik, 2009) menyarankan agar reliabilitas *internal consistency*lah yang paling sesuai untuk digunakan.

Reliabilitas *internal consistency* itu sendiri terdiri dari beberapa jenis analisis untuk jenis tes atau alat ukur yang berbeda. (Reynolds et al., 2009) dan Swerdlik (2009) menyatakan menjelaskan beberapa contoh metode yang berbeda yang dikembangkan untuk mengukur tipe reliabilitas ini, yaitu *split-half reliability*, formula *Spearman-Brown*, formula *Kuder-Richardson*, dan *coefficient alpha*. Dari keempat jenis analisis ini, manakah jenis analisis yang paling tepat untuk digunakan pada skala psikologi? Untuk menjawab pertanyaan ini, para peneliti perlu memahami karakteristik jawaban yang umumnya diberikan oleh responden dalam pengisian sebuah skala psikologi.

Umumnya, jawaban yang diberikan subjek (responden) dalam sebuah skala psikologi berbentuk politomi atau terdapat lebih dari satu jawaban dan tidak ada jawaban salah. Contoh yang paling populer ialah penggunaan skala Likert pada pilihan jawaban atau respons subjek. Ada skala Likert dengan empat pilihan jawaban (contoh: 4 = selalu, 3 =

sering, 2 = jarang, 1 = tidak pernah) dan ada yang terdiri dari lima pilihan jawaban, misalnya 5 = sangat setuju; 4 = setuju; 3 = kurang setuju; 2 = tidak setuju; 1 = sangat tidak setuju). Namun, dari semua jumlah pilihan jawaban yang umum digunakan, Taherdoost (2019) menyarankan untuk menggunakan skala Likert dengan tujuh pilihan jawaban atau dapat menggunakan skala dengan enam pilihan jika ada kebutuhan untuk mengarahkan responden pada satu sisi. Untuk karakteristik jawaban seperti ini, yaitu tidak ada jawaban yang salah (Azwar, 2017b), maka pilihan yang lebih sesuai ialah menggunakan analisis *coefficient alpha*. Analisis ini lebih sesuai digunakan untuk menganalisis jawaban dari skala psikologi karena *coefficient alpha* merupakan analisis yang digunakan pada butir-butir alat ukur yang memiliki multi poin (*multipoint items*) atau yang berbentuk politomi (Reynolds et al., 2009) & (Swerdlik, 2009). Oleh sebab alasan ini, maka analisis *coefficient alpha* merupakan jenis analisis yang paling umum digunakan dalam praktik pengukuran psikologi.

Indeks atau koefisien reliabilitas *coefficient alpha* dapat dihitung secara manual atau menggunakan bantuan program komputer. Untuk menghitung secara manual, peneliti dapat menggunakan rumus berikut (Swerdlik, 2009):

$$r_{\alpha} = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum \sigma_i^2}{\sigma^2} \right)$$

r_{α} = Coefficient alpha
 k = Jumlah butir
 $\sum \sigma_i^2$ = Jumlah varians dari setiap butir
 σ^2 = Varians total skor

Sebagai cara alternatif yang lebih cepat dan efisien, peneliti dapat memilih menggunakan bantuan program komputer untuk melakukan penghitungan. Selain pertimbangan efisiensi, menggunakan bantuan komputer saat melakukan penghitungan ini akan menghilangkan kemungkinan kita melakukan kesalahan penghitungan. Perangkat lunak komputer yang

Reliability Statistics	
Cronbach's Alpha	N of Items
.839	15

Gambar 1. Hasil Analisis Reliabilitas Coefficient Alpha Skala Kesehatan Mental Siswa

dapat digunakan untuk melakukan analisis *coefficient alpha* ialah SPSS atau Statistical Packages for Social Sciences yang dikembangkan oleh IBM. Dalam pengamatan di lapangan, SPSS merupakan perangkat lunak untuk analisis yang paling populer digunakan karena kepraktisan penggunaannya. Contoh hasil analisis reliabilitas *coefficient alpha* menggunakan SPSS dapat dilihat pada Gambar 1 dalam bentuk output SPSS. Dalam contoh ini, peneliti melakukan analisis reliabilitas atas skala kesehatan mental siswa (Gandey, 2023). Oleh peneliti, skala ini digunakan untuk mengumpulkan data kesehatan mental dari siswa SMA di sebuah sekolah. Selanjutnya, untuk melaporkan indeks reliabilitas sebuah skala psikologi yang dihasilkan dari perangkat lunak SPSS dalam teks laporan penelitian

students of Grade XI and their responses were analyzed for reliability. To estimate the reliability of the questionnaires, this study used internal consistency analyses using the coefficient alpha formula. The results showed that the online learning motivation questionnaire was reliable with a coefficient alpha index of .76. The second questionnaire measuring the online learning participation rate was also reliable ($r_{xx} = .92$). Because the coefficient alpha indexes of the two questionnaires were higher than .70, they were considered consistent to measure the constructs.

Gambar 2. Penulisan Hasil Analisis Reliabilitas

dan artikel jurnal, peneliti tidak perlu mengkopi langsung output seperti pada Gambar 1, namun dapat langsung menuliskan hasilnya seperti dalam Gambar 2 (Sinadia & Sondakh, 2023). Oleh sebab terdapat lebih dari satu perangkat lunak komputer yang dapat digunakan untuk analisis reliabilitas dengan tampilan output yang berbeda satu dengan yang lain, maka penulisan yang lebih ringkas bertujuan agar pembaca yang membaca laporan penelitian atau artikel jurnal kita dapat dengan segera memahami tingkat konsistensi dari skala psikologi yang digunakan peneliti.

KESIMPULAN

Dari hasil studi literatur ini, terdapat beberapa kesimpulan mengenai aplikasi dan praktik konstruksi skala psikologi dalam penelitian-penelitian pendidikan. Pertama, validitas dan reliabilitas dari sebuah skala psikologi merupakan persyaratan mutlak yang harus dipenuhi dalam pengembangan jenis instrumen ini. Jika tidak terpenuhi, dapat dipastikan hasil pengukuran yang diperoleh dari alat ukur yang tidak memenuhi kedua syarat ini mengandung banyak error dan tidak dapat diandalkan konsistensinya. Jika ada laporan penelitian dari pengukuran psikologi yang tidak mencantumkan informasi tentang proses pembuktian validasi dan estimasi reliabilitas, kita tidak dapat sepenuhnya mempercayai hasil dari penelitian tersebut. Kedua, pembuatan kisi-kisi (*blueprint*) skala psikologi juga merupakan hal yang bersifat mandatori atau keharusan. Kegagalan seorang peneliti dalam pengukuran psikologi untuk menunjukkan bukti perumusan komponen, indikator, dan subindikator dari skala psikologi yang ia gunakan, menunjukkan buruknya kualitas instrumen yang ia gunakan. Kesimpulan akhir yang dapat ditarik terkait kedua kesimpulan sebelumnya ialah sebuah skala psikologi yang berkualitas baik memiliki informasi yang cukup mengenai proses perumusan kisi-kisi instrumen, proses pembuktian validasi, dan proses estimasi reliabilitas yang dilakukan oleh pengembang instrumen tersebut.

SARAN

Untuk mempermudah penghitungan, teknik analisis validitas dan reliabilitas dapat dilakukan oleh peneliti menggunakan bantuan program komputer. Saat ini sudah terdapat bermacam-macam perangkat komputer yang dapat digunakan untuk melakukan berbagai analisis ini. Peneliti pemula perlu memilih dan mempelajari salah satu dari perangkat lunak yang sesuai kebutuhan untuk digunakan dalam praktik analisis. Gunakanlah perangkat lunak yang mudah dijangkau dari segi biaya dan yang praktis penggunaannya di lapangan.

Oleh sebab kompleksnya analisis validitas konstruk, maka teknik analisis ini umumnya dilakukan oleh para peneliti tingkat menengah dan mahir seperti mereka yang berada pada jenjang pendidikan pasca sarjana. Bagi para mahasiswa tingkat sarjana, biasanya teknik analisis ini belum umum digunakan oleh karena rumitnya analisis yang harus dilakukan. Bagi para peneliti pemula yang ingin menggunakan teknik analisis validitas konstruk, disarankan agar mempelajari dan mendalami teorinya dengan baik atau meminta bantuan ahli yang menguasai teknik analisis ini. Bila peneliti pemula mengalami kesulitan untuk mengembangkan atau mengkonstruksi skala psikologi mereka sendiri, adalah lebih baik memilih untuk menggunakan skala yang telah dikembangkan dengan baik oleh peneliti yang lain. Cara ini dapat mengurangi berbagai kesulitan yang mungkin muncul dalam upaya pengembangan skala psikologi dari peneliti itu sendiri.

REFERENSI

- AERA, APA, & NCME (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Azwar, S. (2014). *Reliabilitas dan validitas* (ed. 4). Pustaka Pelajar.
- Azwar, S. (2017a). *Metode penelitian psikologi* (ed. 2). Pustaka Pelajar.
- Azwar, S. (2017b). *Penyusunan skala psikologi* (ed. 2). Pustaka Pelajar.
- Field, A. (2009). *Discovering statistics using SPSS* (ed. 3). SAGE Publications.
- Gandey, G. G. G. (2023). *The relationship between academic stress and mental health* [Skripsi tidak dipublikasikan]. Universitas Klabat.
- Johnson, R. B., & Christensen, L. (2014). *Educational Research: Quantitative, Qualitative, and Mixed Approaches: Quantitative, Qualitative, and Mixed Approaches* (ed. 5). SAGE.
- Mardapi, D. (2008). *Teknik penyusunan instrumen tes dan nontes*. Mitra Cendikia.
- Pallant, J. (2007). *SPSS: Survival manual* (ed. 3). Mc-Graw Hill.
- Retnawati, H. (2016). *Validitas, reliabilitas, & karakteristik butir: Panduan untuk peneliti, mahasiswa, dan psikometrian*. Parama Publishing.
- Reynolds, C. R., Livingstone, R. B., & Willson, V. (2009). *Measurement and assessment in education* (ed. 2). Pearson Education.
- Sinadia, A. R., & Sondakh, P. M. (2023). Online learning motivation of senior high school students and how it relates to their online learning participation rate. *SUMIKOLAH: Jurnal Ilmu Pendidikan*, 1(1), 31-38.
- Swerdlik, C. (2009). *Psychological testing and assessment: An introduction to tests and measurement* (ed. 7). Mc-Graw Hill.
- Taherdoost, H. (2019). What is the best response scale for survey and questionnaire design; Review of different lengths of rating scale/attitude scale/likert scale. *International Journal of Academic Research in Management*, 8, 1-12.
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In *Handbook of self-regulation: Theory, research, and applications*. Academic Press.