

## VALIDITAS ISI: TAHAP AWAL PENGEMBANGAN KUESIONER

**Hendryadi**

Fakultas Ekonomi, Universitas Islam Attahiriyah, Jakarta, Indonesia

E-mail korespondensi: hendry.basrah@gmail.com

---

### Informasi Artikel

---

Draft awal: 1 Mei 2017  
Revisi : 25 Mei 2017  
Diterima : 20 Juni 2017

---

Kata Kunci:  
*Validitas isi, content validity, CVI*

---

Tipe Artikel : Methodology  
Articles



Diterbitkan oleh Fakultas  
Ekonomi Universitas Islam  
Attahiriyah

---

### ABSTRACT

---

This paper aims to identify the content validity testing techniques in the early stages of developing the questionnaire. Using the literature study, it can be concluded: first, there is no standard technique for accessing the content validity. The researchers can use a qualitative, quantitative, or mix methods. Second, the common qualitative techniques used are through in-depth interviews or FGDs with experts (logical validity) or users (face validity). Third, the common statistical techniques used in testing content validity in the last ten years is the content validity index (CVI) based on the expert agreement level on the item (i-CVI) or S-CVI on the questionnaire as a whole. Other statistical techniques such as Kappa coefficients, ANOVA, and correlation are also commonly used in the initial testing process of the questionnaire. Simulation and case study are discussed in the article.

Artikel ini ditujukan untuk mengidentifikasi teknik pengujian validitas isi pada tahap awal pengembangan kuesioner. Menggunakan metode studi pustaka, dapat diberikan kesimpulan: pertama, tidak ada teknik yang baku dalam pengujian validitas konten. Para peneliti dapat menggunakan pendekatan kualitatif, kuantitatif, atau gabungan keduanya. Kedua, teknik kualitatif yang umum dipergunakan adalah melalui wawancara mendalam atau FGD dengan pakar (*logical validity*) atau pengguna (*face validity*). Ketiga, teknik statistik yang umum dipergunakan dalam pengujian validitas isi dalam sepuluh tahun terakhir adalah content validity index (CVI) yang didasarkan pada tingkat kesepakatan para ahli pada item, atau S-CVI pada kuesioner secara keseluruhan. Teknik statistic lain seperti koefisien Kappa, ANOVA, dan korelasi juga masih lazim digunakan dalam proses pengujian awal kuesioner. Simulasi dan contoh kasus juga dibahas dalam artikel.

---

**Pedoman Sitasi:** Hendryadi, H. (2017). Validitas Isi: Tahap Awal Pengembangan Kuesioner. *Jurnal Riset Manajemen dan Bisnis FE-UNIAT*, 2(2), 169-178

## 1. Pendahuluan

Validitas merupakan isu sentral pada proses pengembangan instrumen, terutama jika digunakan untuk mengukur konsep/konstrak yang masih ambigu, abstrak dan tidak bisa diamati secara langsung. Sebagai contoh, motivasi dapat dinyatakan sebagai proses yang menggerakkan diri seseorang untuk melakukan sesuatu. Konsep ini masih abstrak sehingga diperlukan operasionalisasi lebih lanjut agar dapat diamati dan diukur. Azwar (2012) menyebutkan bahwa para ahli psikometrika telah menetapkan beberapa kriteria penting bagi setiap alat ukur psikologi untuk dinyatakan sebagai alat ukur yang baik, yaitu mampu menghasilkan data dan memberikan informasi yang akurat. Kriteria termaksud antara lain valid, reliable, objektif, standar, ekonomis, dan praktis.

Validitas menggambarkan sejauh alat ukur (tes) benar-benar mengukur apa yang hendak diukur. Menetapkan validitas sebuah test atau instrument test sangat sulit, terutama karena variabel-variabel psikologi biasanya adalah konsep-konsep abstrak, seperti inteligensi, kecemasan, dan kepribadian. Konsep-konsep ini tidak memiliki realitas konkret sehingga eksistensinya harus diinferensi melalui sarana yang tidak langsung (Growth-Marnat, 2011). Ada tiga jenis validitas yaitu *content validity* (validitas isi), *criterion validity* (validitas kriteria), dan *construct validity* (validitas konsep). Dari tiga jenis validitas tersebut, validitas konstrak selalu menjadi yang paling bermasalah dan mendapatkan banyak perdebatan karena melibatkan teori, dan hubungan data dengan teori (Embretson, 2007).

Ketika sebuah skala baru dikembangkan, para peneliti yang mengikuti prosedur pengembangan skala sesuai dengan standar sehingga diharapkan dapat memberikan informasi yang luas mengenai reliabilitas dan validitas skala tersebut. Standar dan prosedur tersebut umumnya dikeluarkan oleh lembaga-lembaga kredibel seperti American Educational Research Association (AERA), American Psychological Association (APA), dan National Council on Measurement in Education (NCME). Meskipun validitas *criterion-related validity* dan *construct validity* dianggap sangat penting, informasi tentang validitas konten juga dipandang perlu dalam menarik kesimpulan tentang kualitas skala (Polit dan Beck, 2006).

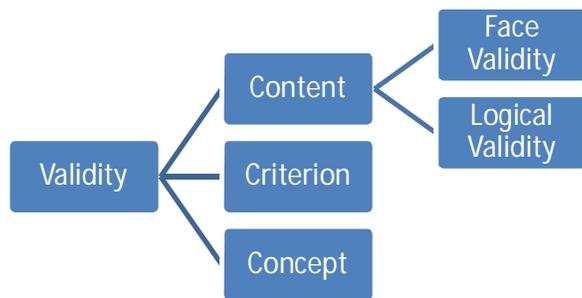
Artikel ini secara khusus akan membahas mengenai isu validitas pertama yaitu validitas isi yang lebih sering dipergunakan pada tahap-tahap awal pengembangan instrumen penelitian. Terlepas dari berbagai perdebatan mengenai kesatuan konsep validitas seperti yang disampaikan oleh Lissitz dan Samuelsen (2007), evaluasi empiris mengenai penggunaan uji validitas isi pada berbagai publikasi ilmiah terkait dengan pengembangan skala penelitian dapat memberikan gambaran umum mengenai tahapan awal pengembangan skala yang berlaku umum sekarang ini.

## 2. Kajian Pustaka

### 2.1. Tentang Validitas

Validitas mengacu pada aspek ketepatan dan kecermatan hasil pengukuran. Pengukuran sendiri dilakukan untuk mengetahui seberapa banyak aspek (dalam arti kuantitatif) suatu aspek psikologis terdapat dalam diri seseorang, yang dinyatakan oleh skornya pada instrument pengukur yang bersangkutan. Dalam konteks Teori Skor-murni Klasik, Azwar (2012) lebih lanjut menjelaskan bahwa makna validitas dapat dinyatakan sebagai sejauhmana besaran skor-tampak X mampu mendekati besaran skor-murni T. Semakin skor-tampak mendekati skor-murni berarti semakin tinggi validitas dan sebaliknya sebaliknya, semakin rendah validitas hasil pengukuran berarti semakin besar perbedaan skor-tampak dari skor-murni.

Jika reliabilitas menyangkut isu-isu konsistensi, maka validitas mengakses akurasi test. Validitas adalah sejauh alat ukur (tes) benar-benar menggambarkan apa yang hendak diukur. Menetapkan validitas sebuah test atau instrument test sangat sulit, terutama karena variabel-variabel psikologi biasanya adalah konsep-konsep abstrak, seperti inteligensi, kecemasan, dan kepribadian. Konsep-konsep ini tidak memiliki realitas konkret sehingga eksistensinya harus diinferensi melalui sarana yang tidak langsung (Growth-Marnat, 2011). Selain itu, konseptualisasi dan penelitian tentang konstrak sering mengalami perubahan seiring berjalannya waktu, sehingga validasi sebuah alat ukur harus terus-menerus disempurnakan. Ada tiga jenis validitas yaitu *content validity* (validitas isi), *criterion validity* (validitas kriteria), dan *construct validity* (validitas konsep).



Gambar 1. Jenis Uji Validitas

Validitas bukan semata-mata sebagai sebuah tes, sebaliknya, ini mengacu pada penggunaan tes untuk tujuan tertentu. Dengan demikian, validitas merupakan upaya peneliti untuk mengevaluasi kegunaan dan kelayakan tes untuk tujuan tertentu yang memerlukan banyak sumber bukti. Hal ini diperlukan jika penggunaan tes harus dipertahankan untuk tujuan tertentu, sehingga bukti yang memadai dapat diajukan untuk mempertahankan penggunaan tes untuk tujuan itu. Selain itu, evaluasi validitas bukan kejadian statis satu kali, namun merupakan proses yang terus menerus (Sireci, 2007).

#### **Validitas Isi (*Content Validity*)**

Dalam Standar untuk Tes Pendidikan dan Psikologi (American Educational Research Association, American Psychological Association & National Council of Measurement in Education, 1985:9) "bukti validitas terkait konten" didefinisikan sebagai "bukti yang menunjukkan sejauh mana isi tes sesuai dengan tujuan yang dimaksudkannya. Bukti tersebut digunakan untuk menetapkan bahwa pengujian tersebut mencakup contoh yang representatif atau kritis dari domain konten yang relevan dan isinya tidak termasuk konten di luar domain tersebut".

Validitas isi merupakan validitas yang diestimasi lewat pengujian terhadap kelayakan atau relevansi isi tes melalui analisis rasional oleh panel yang berkompeten atau melalui expert judgement (penilaian ahli). Validitas isi atau *content validity* memastikan bahwa pengukuran memasukkan sekumpulan item yang memadai dan mewakili yang mengungkap konsep. Semakin item skala mencerminkan kawasan atau keseluruhan konsep yang diukur, semakin besar validitas isi. Atau dengan kata lain, validitas isi merupakan fungsi seberapa baik dimensi dan elemen sebuah konsep yang telah digambarkan (Sekaran, 2006).

Mengukur dan melaporkan validitas konten instrumen adalah penting, karena jenis validitas ini juga dapat membantu memastikan validitas konstruk dan memberi kepercayaan kepada pembaca dan peneliti tentang instrument karena melibatkan pakar-pakar untuk memeriksa kelayakan instrument dari sisi konsep dan operasionalisasi. Konseptualisasi dan operasionalisasi merupakan tahap awal pengembangan kuesioner (Growth-Marnat, 2010), dan terkadang meskipun peneliti sudah mengikuti langkah-langkah itu dengan taat, kadang-kadang tetap sulit untuk menentukan apa yang sesungguhnya diukur oleh tes tersebut. Beberapa istilah lain dari validitas isi antara lain: *content related validity*, *intrinsic validity*, *relevance validity*, *representative validity* dan *logical* atau *sampling validity*.

#### **Validitas Berdasarkan Kriteria (*criterion-related validity*)**

Validitas berdasarkan kriteria atau *criterion-related validity* merupakan sebuah ukuran validitas yang ditentukan dengan cara membandingkan skor-skor tes dengan kinerja tertentu pada sebuah ukuran luar. Ukuran luar ini seharusnya memiliki hubungan teoritis dengan variabel yang diukur oleh tes itu. Misalnya, tes inteligensi mungkin berkorelasi dengan rata-rata nilai akademis,

Validitas criteria (*criterion-related validity*) terpenuhi jika pengukuran membedakan individu menurut suatu criteria yang diharapkan diprediksi. Hal tersebut bisa dilakukan dengan menghasilkan validitas konkuren (*concurrent validity*) atau validitas predictive (*predictive validity*). Validitas konkuren dihasilkan jika skala membedakan individu yang diketahui berbeda, yaitu mereka harus menghasilkan skor yang berbeda pada instrument, sedangkan validitas predictive menunjukkan kemampuan instrument pengukuran untuk membedakan orang dengan referensi pada suatu criteria masa depan (Sekaran, 2006). Dengan demikian, perbedaan antara *concurrent validity* dengan *predictive validity* adalah waktu pengujian, dimana *concurrent validity* diambil dalam waktu yang sama (atau kurang lebih sama), sedangkan *predictive validity* dilakukan beberapa saat (dalam periode waktu tertentu) setelah terlebih dahulu skor hasil tes diperoleh.

### **Validitas Konsep (*Construct Validity*)**

Sebuah tes telah validitas konstruk jika menunjukkan hubungan antara nilai tes dan prediksi sifat teoritis. Groth-Marnat (2010) memberikan penjelasan bahwa "metode *Construct Validity* dikembangkan sebagian untuk mengoreksi ketidak-adekuatan dan kesulitan yang dialami dengan pendekatan *content* dan pendekatan *criterion*. Bentuk-bentuk awal validitas isi terlalu banyak mengandalkan pada judgement subjektif, sementara validitas *criterion* terlalu restriktif dalam bekerja dengan ranah-ranah atau struktur konstruk-konstruk yang diukur. Validitas *criterion* memiliki kesulitan lain dalam arti bahwa sering kali tidak ada kesepakatan dalam menetapkan kriteria luar yang adekuat".

Pendekatan dasar validitas konstruk adalah mengakses sejauh mana test yang dimaksud mengukur sebuah konstruk teoritis atau ciri-sifat. Asesmen ini melibatkan 3 langkah umum, yaitu : pertama, konstruktor tes harus melakukan analisis yang diteliti terhadap konsep. Kedua, mempertimbangkan bagaimana hubungan sifat-ciri itu dengan variabel lain. Ketiga, perancang tes perlu menguji dulu apakah hubungan-hubungan dihipotesiskan benar-benar ada (Foster & Cone, 1995, dikutip oleh Groth-Marnat, 2010).

### **2.2. Jenis Validitas Isi**

Validitas isi dilakukan untuk memastikan apakah isi kuesioner sudah sesuai dan relevan dengan tujuan study. Validitas isi menunjukkan isi mencerminkan rangkaian lengkap atribut yang diteliti dan biasanya dilakukan oleh tujuh atau lebih ahli (DeVon et al 2007). Perkiraan validitas isi dari tes diperoleh dengan menyeluruh dan sistematis dalam memeriksa item tes untuk menentukan sejauh mana mereka mencerminkan dan tidak mencerminkan domain konten (Kowsalya, Venkat Lakshmi, dan Suresh, 2012)

Validitas tampak/muka (*face validity*) merupakan validitas isi yang paling dasar dan sangat minimum. Validitas isi menunjukkan bahwa item-item yang dimaksudkan untuk mengukur sebuah konsep, memberikan kesan mampu mengungkap konsep yang hendak di ukur (Sekaran, 2006). Tidak berbeda dengan penjelasan sebelumnya, Groth-Marnat, (2010) menjelaskan bahwa validitas isi (*content validity*) dengan validitas muka (*face validity*) memiliki perbedaan dan tidak sinonim. Validitas isi menyangkut judgement yang dibuat oleh para ahli, sedangkan validitas muka/tampang menyangkut *judgement* dari pengguna test. Sejalan dengan itu, Gregory (1992) yang dikutip Azwar (2012) menjelaskan bahwa validitas tampak hanya sekedar tahap penerimaan orang pada umumnya terhadap fungsi pengukuran tes, serta tidak berhubungan dengan statistic validitas seperti koefisien atau indeks.

Analisis lanjutan setelah validitas tampak adalah melalui validitas logis yaitu prosedur penilaian kelayakan isi item melalui penilaian yang bersifat kualitatif oleh panel ahli. Prosedur ini selanjutnya menghasilkan validitas logis atau merupakan tinggi rendahnya kesepakatan di antara para ahli yang menilai kelayakan suatu skala pengukuran (Azwar, 2012).

Lawshe (1975) mengusulkan rasio validitas isi (CVR) untuk mengukur derajat kesepakatan para ahli dari satu item dan yang dapat mengekspresikan tingkat validitas konten melalui indicators tunggal yang berkisar dari -1 sampai 1. Pendekatan lain adalah koefisien validitas isi dan reliabilitas koefisien homogenitas diusulkan oleh Aiken (1980, 1985), yang dapat digunakan untuk mengukur peringkat validitas setiap item (V value). Dua teknik yang digunakan dalam analisis ini adalah melalui koefisien validitas isi Aiken's V dan Rasio validitas isi - Lawshe's CVR.

### Koefisien Validitas Isi – Aiken's V

Aiken (1985) merumuskan formula Aiken's V untuk menghitung *content-validity coefficient* yang didasarkan pada hasil penilaian dari panel ahli sebanyak n orang terhadap suatu item dari segi sejauh mana item tersebut mewakili konstruk yang diukur.

Formula yang diajukan oleh Aiken adalah sebagai berikut (dalam Azwar, 2012)

$$V = \sum s / [n(C-1)]$$

$$S = r - lo$$

Lo = angka penilaian terendah (misalnya 1)

C = angka penilaian tertinggi (misalnya 4)

R = angka yang diberikan oleh penilai

Tabel 1. Contoh Validitas Isi Aiken's V

Penilai	Item 1	
	Nilai (R)	S = R - Lo
1	3	3 - 1 = 2
2	4	4 - 1 = 3
3	4	4 - 1 = 3
4	3	3 - 1 = 2
5	3	3 - 1 = 2
6	4	4 - 1 = 3
7	4	4 - 1 = 3
	$\sum s$	18
	V	0.857

Nilai S untuk penilai 1 diperoleh dari skor penilai 3 dikurangi skor terendah (Lo), sehingga 3 - 1 = 2, dan seterusnya. Nilai  $\sum s$  adalah penjumlahan dari skor S yaitu 2 + 1 + 3 + 2 + 2 + 3 + 3 = 18. Dengan demikian, nilai V dapat dihitung sebagai berikut :

$$V = \sum s / [n(C-1)]$$

$$V = 18 / [7(4-1)]$$

$$V = 0.857$$

Nilai koefisien Aiken's V berkisar antara 0 - 1. Koefisien sebesar 0.857 ini sudah dapat dianggap memiliki validitas isi yang memadai.

### Koefisien Validitas Isi – Lawshe's CVR

Lawshe's CVR (content validity ratio) merupakan salah satu metode yang digunakan secara luas untuk mengukur validitas isi. Teknik ini dikembangkan oleh Lawshe (1975). Pendekatan ini pada dasarnya adalah sebuah metode untuk mengukur kesepakatan di antara penilai atau hakim tentang pentingnya item tertentu.

Lawshe (1975) mengusulkan bahwa setiap penilai / *subject matter experts* (SME) yang terdiri dari panel ahli untuk menjawab pertanyaan untuk setiap item dengan tiga pilihan jawaban yaitu (1) esensial, (2) berguna tapi tidak esensial, (3) tidak diperlukan. Menurut Lawshe, jika lebih dari setengah panelis menunjukkan bahwa item penting/esensial, maka item tersebut memiliki setidaknya validitas isi yang cukup. Formula yang diajukan oleh Lawshe adalah :

$CVR = (n_e - N/2) / (N/2)$ , dimana CVR adalah content validity ratio,  $n_e$  adalah jumlah anggota panelis yang menjawab “penting”, N adalah jumlah total panelis.

Misalnya, Seorang peneliti ingin menguji validitas isi dari sebuah skala yang terdiri dari 1 item. Sebanyak 12 orang panel ahli dijadikan penilai dengan memilih 3 pilihan jawaban yaitu “penting”, “sesuai, tidak penting” dan “tidak berguna”.

Dari 12 orang panel ahli, 10 orang menyatakan item tersebut “penting”, 1 orang menyatakan “sesuai, tidak penting”, dan 1 orang menyatakan “tidak berguna (Lihat tabel 2). Dari data ini kemudian dapat dihitung CVR sebagai berikut :

$$CVR = ((10 - (12/2)) / (12/2))$$

$$CVR = 0.667$$

Formula ini menghasilkan nilai-nilai yang berkisar dari +1 sampai -1, nilai positif menunjukkan bahwa setidaknya setengah panelis menilai item sebagai penting/esensial. Semakin lebih besar CVR dari 0, maka semakin “penting” dan semakin tinggi validitas isinya. Dari contoh di atas diperoleh nilai CVR sebesar 0.667 yang menunjukkan bahwa item yang digunakan sudah memenuhi validitas isi yang baik dan dianggap penting lebih dari separuh ahli yang digunakan sebagai penilai.

Tabel 2. Data Simulasi Validitas Isi Lawshe's CVR

Penilai (SME)	Penilaian
1	Penting
2	Penting
3	Penting
4	Tidak Berguna
5	Penting
6	Penting
7	Sesuai, Tidak Penting
8	Penting
9	Penting
10	Penting
11	Penting
12	Penting
Jawaban Penting = 10	

### Content Validity Index (CVI)

Salah satu pendekatan yang juga direkomendasikan untuk mengevaluasi validitas isi adalah *content validity index* (CVI). Pendekatan ini, sama dengan dua pendekatan sebelumnya, juga melibatkan tim ahli untuk menentukan apakah setiap item dalam skala susah sesuai atau relevan dengan konstraknya, menghitung persentase item yang dianggap relevan untuk setiap pakar, dan kemudian mengambil rata-rata persentase di antara pakar. Sebagai contoh, misalnya dengan tiga ahli, jika Expert 1 memberikan nilai kesesuaian sekumpulan item dengan kostruknya sebesar 90%, Expert 2 memberikan nilai kesesuaian 100%, dan Expert 3 memberi nilai 90%, maka nilai indeks ini adalah 93%. Ini disebut sebagai persentase kongruensi rata-rata (*average congruency percentage* - ACP). Nilai ACP disarankan lebih besar dari 90 sebagai syarat diterimanya sebuah instrumen (Popham, 1978, Waltz et al., 2005)

*Content validity index (CVI)* merupakan salah satu teknik yang paling banyak digunakan dalam riset keperawatan (Polit dan Beck, 2006). Teknik ini dikembangkan oleh Martuza (1977), seorang spesialis pendidikan. Meski demikian, CVI telah memiliki banyak kritik, dan kemudian Lynn (1986), dalam penelitiannya menghitung dua jenis CVI. Tipe pertama melibatkan validitas isi item individual (i-CVI) dan yang kedua melibatkan validitas konten dari skala keseluruhan (s-CVI).

Lynn (1986) menganjurkan minimal menggunakan tiga ahli, namun mengindikasikan bahwa lebih dari 10 mungkin tidak perlu. Skala pengukuran yang disarankan adalah skala ordinal 4 titik untuk poin untuk menghindari titik tengah netral dan ambivalen. Beberapa label yang sering digunakan: 1 = tidak relevan, 2 = agak relevan, 3 = cukup relevan, 4 = sangat relevan. Kemudian, untuk setiap item, I-CVI dihitung sebagai jumlah ahli yang memberikan penilaian baik yaitu 3 atau 4 (dengan demikian dikotomisasi skala ordinal menjadi relevan = 1 dan tidak relevan = 0), dibagi dengan jumlah total ahli. Misalnya, item yang dinilai cukup atau sangat relevan oleh empat dari lima penilai akan memiliki I-CVI sebesar 0,80 (Polit dan Beck, 2006).

Tabel 3. Data Simulasi CVI

Item	Expert 1	Expert 2	Expert 3	Jumlah Kesetujuan	I-CVI
1	0	1	1	2	2/3 = 0.67
2	1	0	1	2	2/3 = 0.67
3	1	1	0	2	2/3 = 0.67
4	1	1	1	3	3/3 = 1.00
5	1	1	1	3	3/3 = 1.00
6	1	1	1	3	3/3 = 1.00
7	1	1	1	3	3/3 = 1.00
8	1	1	1	3	3/3 = 1.00
9	1	1	1	3	3/3 = 1.00
10	1	1	1	3	3/3 = 1.00
$\Sigma$	9	9	9	Mean I-CVI	0.90
Proposi Relevan	0.90	0.90	0.90		

Mean I-CVI, item-level content validity index rata-rata = 0.90

Proporsi rata-rata item dinilai relevan di tiga ahli = 0,90

Nilai I-CVI harus 1,00 bila ada lima atau lebih sedikit penilai. Bila ada enam atau lebih penilai, standarnya bisa lebih longgar, tapi Lynn (1986) merekomendasikan I-CVI tidak lebih rendah dari 0,78.

### **Interrater reliability (Kappa Statistic)**

Teknik kuantitatif ke empat yang juga banyak digunakan adalah Kappa Statistik. Teknik ini digunakan beberapa peneliti seperti Wynd, Schmidt, dan Schaefer (2003) yang menggunakan CVI dan kappa multi-rater dalam validasi konten skala yang dikembangkannya. Mereka berpendapat bahwa statistik kappa merupakan suplemen penting, namun bukan pengganti CVI. Kappa memberikan informasi tentang tingkat kesepakatan di luar kemungkinan (Polit et al., 2007).

Kappa statistic atau *interrater reliability* adalah ukuran yang digunakan untuk menguji kesepakatan antara dua orang (penilai / pengamat) pada variabel kategoris. Jika lebih dari 2 penilai, maka teknik multi-rater Kappa dapat digunakan. Ukuran statistik reliabilitas interrater adalah Kappa Cohen yang umumnya berkisar antara 0 sampai 1.0 (walaupun angka negatifnya dimungkinkan) di mana angka besar berarti keandalan yang lebih baik, nilai mendekati atau kurang dari nol menunjukkan bahwa kesepakatan disebabkan oleh kebetulan saja. Analisis Kappa dapat dilakukan menggunakan alat bantu SPSS pada menu Tabulasi Silang (*Crosstabulation*). Landis dan Koch (1977) memberikan pedoman penilaian pada Kappa statistic sebagai berikut :

Tabel 4. Evaluasi Nilai Statistik Kappa

Kappa	Interpretation
< 0	Poor agreement
0.0 – 0.20	Slight agreement
0.21 – 0.40	Fair agreement
0.41 – 0.60	Moderate agreement
0.61 – 0.80	Substantial agreement
0.81 – 1.00	Almost perfect agreement

Sumber : Landis, J. R., Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33:159-174

### 3. Metode

Pendekatan yang digunakan dalam artikel ini adalah studi literatur untuk memperoleh penjelasan komprehensif mengenai berbagai teknik validitas konten beserta berbagai keterbatasannya. Data dikumpulkan melalui pencarian artikel di google scholar dengan menggunakan keyword “content validity” dan “scale development”. Meskipun mesin pencari menampilkan banyak artikel yang relevan dengan kata kunci yang digunakan, namun untuk mempermudah analisa, artikel yang akan direview hanyalah artikel bertipe *open access* atau dapat diunduh secara bebas.

Artikel dipilih didasarkan pada jumlah sitasi yang terekam di google scholar, disini, minimal sitasi 150 dianggap layak untuk dijadikan acuan dalam pengambilan kesimpulan dari topik validitas isi. Berikut ini ditampilkan artikel yang dipilih dalam pembahasan :

Tabel 5. Artikel sebagai sumber pembahasan

No	Penulis	Jumlah Sitasi	Informasi Artikel
1	Polit dan Beck (2006)	1284	Polit, D. F., & Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? Critique and recommendations. <i>Research in nursing &amp; health</i> , 29(5), 489-497
2	Polit, Beck, & Owen (2007)	907	Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. <i>Research in nursing &amp; health</i> , 30(4), 459-467.
3	Wynd, Schmidt, & Schaefer (2003)	347	Wynd, C. A., Schmidt, B., & Schaefer, M. A. (2003). Two quantitative approaches for estimating content validity. <i>Western Journal of Nursing Research</i> , 25(5), 508-518.
4	Hinkin, Tracey, & Enz, (1997).	186	Hinkin, T. R., Tracey, J. B., & Enz, C. A. (1997). Scale construction: Developing reliable and valid measurement instruments. <i>Journal of Hospitality &amp; Tourism Research</i> , 21(1), 100-120.
5	Hinkin, (1995)	2001	Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. <i>Journal of management</i> , 21(5), 967-988.

Keterangan : jumlah sitasi didasarkan pada indeks google scholar tanggal 1 Juni 2017

#### 4. Pembahasan

Hitkin (1995) memberikan kesimpulan bahwa validitas isi diperlukan pada tahap pengembangan skala. Salah satu metode yang umum mengharuskan responden untuk mengelompokkan atau mengurutkan item berdasarkan kesamaannya untuk membangun definisi konsep. Hal ini dapat dilakukan dengan menggunakan pakar untuk mengevaluasi kesesuaian item dengan konsep yang diukur. Indeks kesepakatan yang dapat diterima harus ditentukan sebelum administrasi item dan definisi.

Metode yang lebih baru dikembangkan untuk melakukan penilaian konten menggunakan teknik sortasi dan analisis faktor untuk menilai secara kuantitatif kecukupan konten dari serangkaian item yang baru dikembangkan. Responden diminta untuk menilai sejauh mana item sesuai dengan definisi konstruk. Tanggapan kemudian dianalisis menggunakan analisis faktor dan item dinyatakan tepat dapat dipertahankan untuk administrasi selanjutnya ke sampel tambahan. Metode ketiga, berdasarkan analisis teknik varians. ini. Teknik ini sangat sederhana dan mudah dan memungkinkan uji statistik untuk menguji kelayakan konten. Hal ini dapat dilakukan dengan sampel yang relatif kecil, sehingga dapat meminimalkan biaya maupun waktu pengumpulannya. Meski demikian, Hitkin (1997) mencatat bahwa tak satu pun dari teknik ini akan menjamin bahwa skala memiliki konten yang valid, namun akan memberikan bukti bahwa item tersebut merupakan ukuran yang wajar dari konstruksi di bawah pengujian dan mengurangi kebutuhan akan modifikasi skala berikutnya.

Polit dan Beck (2006) berdasarkan hasil studinya memberikan kesimpulan bahwa kejelasan tentang validasi konten dalam studi pengembangan skala diperlukan. Mereka menawarkan beberapa rekomendasi yang akan memperbaiki komunikasi tentang validitas konten yang digunakan peneliti mengenai penggunaan CVI untuk mengukur kesepakatan tentang relevansi item. Meski demikian, Polit dan Beck (2006) dan Polit et al., (2007) memberikan perbedaan pada dua teknik kuantitatif untuk mengevaluasi validitas konten yaitu I-CVI dan S-CVI. I-CVI lebih ditujukan untuk mengukur kesepakatan ahli di tingkat item, sedangkan S-CVI digunakan untuk mengukur kesepakatan ahli di tingkat kuesioner secara keseluruhan.

Wynd, Schmidt, & Schaefer (2003) memberikan kesimpulan bahwa validitas isi instrumen sering ditemukan melalui tinjauan ahli kualitatif, namun analisis kuantitatif terhadap kesepakatan peninjau juga dianjurkan dalam literatur. Dua pendekatan kuantitatif digunakan oleh mereka untuk mengembangkan skala Risiko Osteoporosis (ORAT) yaitu Indeks Validitas Isi (CVI) dan koefisien kapa multirater.

Rubio et al., (2003) dalam artikel "*Objectifying content validity: conducting a content validity study in social work research*" memberikan simulasi bagaimana menguji validitas isi. Meskipun validitas isi bersifat subjektif, menggunakan beberapa teknik seperti *interated agreement*, *content validity index (CVI)*, dan *factorial validity index* dapat meningkatkan objektivitas hasil uji. Seperti halnya Hitkin (1997), mereka juga berkesimpulan bahwa pengujian validitas bersifat terus menerus dan tidak cukup menggunakan salah satu teknik dalam pengujiannya.

Berdasarkan uraian di atas dapat dinyatakan bahwa validitas konten masih penting digunakan sebagai langkah awal pengembangan instrumen. Meskipun tidak ada kesepakatan ahli mengenai teknik yang digunakan, namun berbagai riset terkini umumnya menggunakan gabungan metode kualitatif (interview, FGD, pendapat ahli) untuk memperoleh validitas konten yang memadai.

#### 5. Kesimpulan dan Rekomendasi

Berdasarkan pembahasan, maka dapat disimpulkan beberapa poin: pertama, tidak ada teknik yang baku dalam pengujian validitas konten. Para peneliti dapat menggunakan pendekatan kualitatif, kuantitatif, atau gabungan keduanya. Kedua, teknik kualitatif yang umum dipergunakan adalah melalui wawancara mendalam atau FGD dengan pakar (*logical validity*) atau pengguna (*face validity*). Ketiga, teknik statistik yang umum dipergunakan dalam pengujian validitas isi dalam sepuluh tahun terakhir adalah content validity index (CVI) yang didasarkan pada tingkat

kesepakatan para ahli pada item, atau S-CVI pada kuesioner secara keseluruhan. Teknik statistic lain seperti koefisien Kappa, ANOVA, dan korelasi juga masih lazim digunakan dalam proses pengujian awal kuesioner.

### Daftar Pustaka

- Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and psychological measurement*, 40(4), 955-959.
- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and psychological measurement*, 45(1), 131-142.
- Azwar, S. (2012). *Reliabilitas dan Validitas*. Edisi 4. Yogyakarta : Pustaka Pelajar.
- DeVon, H. A., et. al. (2007). psychometric Toolbox for testing Validity and Reliability. *Journal of Nursing scholarship*, 39 (2), 155-164.A
- Embretson, S.E. (2007). Construct Validity: A Universal Validity System or Just Another Test Evaluation Procedure?. *Educational Researcher*, Vol. 36, No. 8: 449-455
- Furr, M. (2011). *Scale construction and psychometrics for social and personality psychology*. SAGE Publications Ltd.
- Growth – Marnat, G. (2010). *Handbook of Psychological Assessment*. Terj. Soetjipto, H.P & Soetjipto, S.M. Yogyakarta : Pustaka Pelajar
- Kowsalya, D. N., Lakshmi, H. V., & Suresh, K. P. (2012). Development and Validation of a Scale to assess Emotional Maturity in Mild Intellectually Disabled Children. *Language in India*, 12(6).Landis, J. R., Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33:159-174
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel psychology*, 28(4), 563-575.
- Lissitz, R. W., & Samuelson, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational researcher*, 36(8), 437-448.
- Polit, D. F., & Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? Critique and recommendations. *Research in nursing & health*, 29(5), 489-497.
- Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in nursing & health*, 30(4), 459-467.
- Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social work research*, 27(2), 94-104.
- Sekaran, U. (2006). *Metode Penelitian Bisnis*. Jakarta : Salemba Empat
- Suryani, & Hendryadi. (2015). *Metode Riset Kuantitatif Teori dan Aplikasi Pada Penelitian Bidang Manajemen dan Ekonomi Islam*. Jakarta: Prenadamedia Group.
- Wynd, C. A., Schmidt, B., & Schaefer, M. A. (2003). Two quantitative approaches for estimating content validity. *Western Journal of Nursing Research*, 25(5), 508-518.
- Yalow, E. S., & Popham, W. J. (1983). Content validity at the crossroads. *Educational Researcher*, 12(8), 10-21.