

LECTURE NOTED

[Entrepreneur & Management Research] Week 09# – Instrument Development: Validity and Reliability

1 — REVIEW OF MEETING 8

Data Collection in Entrepreneurial Contexts

- Data collection is where theory meets reality (e.g., entrepreneurial resilience only becomes researchable when it is connected to actual experiences, actions, or decisions).
- Different methods produce different kinds of evidence (e.g., interviews reveal interpretation, while platform data reveal recorded behavior).
- Entrepreneurial phenomena are often dynamic and processual (e.g., founders may change strategy or reinterpret failure during the study).
- Good data collection depends on fit between the question and the source of evidence (e.g., crowdfunding legitimacy may be better studied through campaign texts than through memory-based recall alone).

Penjelasan

Pada pertemuan sebelumnya, kita telah menegaskan bahwa data collection bukan sekadar pekerjaan lapangan, tetapi titik ketika teori mulai diuji terhadap kenyataan. Konsep seperti resilience, legitimacy, entrepreneurial identity, atau strategic adaptation tidak otomatis menjadi pengetahuan hanya karena sudah didefinisikan di literatur. Ia baru menjadi pengetahuan ketika peneliti mampu menghubungkannya dengan bukti empiris yang benar-benar dapat diamati, ditanyakan, atau ditelusuri. Itulah sebabnya pengumpulan data selalu merupakan tahap epistemologis, bukan sekadar operasional. Textbook metodologi bisnis juga menempatkan tahap ini sebagai jembatan antara research design dan empirical evidence.

Kita juga melihat bahwa metode yang berbeda membuka lapisan realitas yang berbeda. Wawancara memberi akses pada makna, observasi memberi akses pada praktik, dokumen memberi akses pada representasi, dan data digital memberi akses pada perilaku yang terekam. Artinya, ketika peneliti memilih metode pengumpulan data, ia sebenarnya sedang memilih bentuk kebenaran apa yang paling mungkin muncul. Karena itu, kualitas penelitian tidak hanya ditentukan oleh banyaknya data, tetapi oleh kecocokan antara pertanyaan penelitian dan bentuk evidensi yang dipilih.

Dari sini kita masuk ke persoalan yang lebih dalam. Setelah data dikumpulkan, pertanyaan berikutnya bukan lagi hanya “apakah saya punya data,” tetapi “apakah alat yang saya gunakan benar-benar menangkap konstruk yang saya maksud, dan apakah hasilnya cukup konsisten untuk dipercaya?” Di titik inilah pembahasan tentang instrument development, validity, and reliability menjadi sangat penting. Artikel-artikel scale development di jurnal Scopus bereputasi menunjukkan bahwa pengembangan instrumen justru sering menjadi kontribusi ilmiah tersendiri, bukan sekadar pelengkap analisis.

2 — INTRODUCTION

Why Instrument Development Matters

- A research instrument connects abstract constructs to measurable evidence (e.g., founder social identity must be translated into observable indicators before it can be analyzed).
- Poor instruments weaken strong theories (e.g., a vague scale of customer value will produce unclear findings even with a large sample).
- Instrument development is conceptual before it is statistical (e.g., item wording depends on how the construct is theoretically defined).
- Validity and reliability matter because measurement is the foundation of inference (e.g., inconsistent or invalid measures cannot support strong conclusions).

Penjelasan

Instrumen penelitian adalah jembatan antara konsep abstrak dan dunia empiris. Peneliti tidak pernah secara langsung “mengukur” social identity, digital orientation, work-life fulfilment, masstige value, atau entrepreneurial hustle. Yang diukur adalah indikator-indikator yang dianggap mewakili konstruk itu. Karena itu, kualitas kesimpulan penelitian sangat bergantung pada kualitas jembatan pengukuran tersebut. Jika jembatan itu lemah, maka bahkan teori yang sangat baik pun akan menghasilkan temuan yang kabur atau menyesatkan.

Masalahnya, banyak peneliti pemula memperlakukan instrumen sebagai daftar pertanyaan yang disusun setelah variabel ditentukan. Cara berpikir ini terlalu dangkal. Instrument development bukan sekadar menulis item, tetapi menerjemahkan makna konseptual ke dalam bentuk empiris yang dapat dipertanggungjawabkan. Karena itu, sebelum statistik digunakan, peneliti harus terlebih dahulu menjawab pertanyaan yang lebih penting: apa sebenarnya konstruk ini, apa dimensinya, dan apa batas maknanya? Textbook metodologi menekankan bahwa pengukuran yang baik selalu dimulai dari definisi konseptual yang jelas.

Di sinilah validitas dan reliabilitas memperoleh peran sentral. Validitas berbicara tentang apakah instrumen sungguh-sungguh menangkap konstruk yang dimaksud. Reliabilitas berbicara tentang apakah pengukuran itu cukup konsisten untuk dipercaya. Tanpa keduanya, peneliti mungkin memiliki data yang banyak, tetapi tetap tidak memiliki dasar yang kuat untuk membuat inferensi. Dalam penelitian yang serius, alat ukur bukan hanya sarana teknis; ia adalah fondasi kredibilitas temuan.

3 — WHAT IS A RESEARCH INSTRUMENT?

From Abstract Idea to Empirical Tool

- An instrument is a structured way of capturing a construct (e.g., a founder’s social identity can be captured through items that reflect role centrality and self-definition).
- Instruments may take the form of scales, checklists, indexes, or coded protocols (e.g., an empowerment construct can be operationalized through multidimensional scale items).

- A good instrument does not merely ask questions; it operationalizes theory (e.g., hustle is not measured by generic “hard work” items, but by theory-based urgent and unorthodox action indicators).
- An instrument must be interpretable, not merely answerable (e.g., respondents may answer an item, but the question is whether the answer truly represents the construct).

Penjelasan

Instrumen penelitian pada dasarnya adalah alat yang membuat konstruk abstrak bisa “hadir” dalam bentuk data. Tetapi instrumen yang baik bukan hanya kumpulan pertanyaan. Ia adalah hasil dari proses pemikiran yang menghubungkan definisi teoretis dengan indikator empiris. Artikel *measuring the social identity of entrepreneurs* menunjukkan hal ini dengan sangat jelas: identitas sosial entrepreneur tidak dibiarkan sebagai gagasan kabur, tetapi diterjemahkan menjadi item-item yang memungkinkan peneliti mengukur bagaimana founder melihat dirinya dan perannya dalam penciptaan venture.

Bentuk instrumen pun bisa beragam. Dalam beberapa studi, instrumen berbentuk skala psikometrik multidimensi. Dalam studi lain, ia bisa berupa indeks, protokol coding, atau rubrik penilaian. Artikel *frontline employee empowerment* misalnya mengembangkan empowerment sebagai second-order construct dengan beberapa dimensi yang berbeda, sehingga instrumen tidak cukup hanya dengan satu atau dua pertanyaan umum. Ini menunjukkan bahwa bentuk instrumen harus mengikuti kompleksitas konstruk.

Karena itu, instrumen yang baik tidak bertanya secara acak. Ia mengoperasionalkan teori. Artikel *entrepreneurial hustle* memperlihatkan bahwa hustle tidak bisa diukur dengan item generik tentang kerja keras atau semangat kerja. Konstruk itu harus dibangun dari logika tindakan mendesak dan tidak ortodoks di bawah ketidakpastian. Ini memberi pelajaran penting: ketika instrumen lepas dari teori, ia mudah terjebak menjadi kumpulan item yang tampak masuk akal tetapi sebenarnya tidak presisi.

Pada akhirnya, instrumen tidak cukup hanya bisa dijawab. Ia harus bisa ditafsirkan. Jawaban responden hanya berguna bila peneliti bisa menjelaskan apa arti skor itu dalam kaitannya dengan konstruk. Inilah yang membedakan instrumen akademik dari sekadar daftar pertanyaan biasa.

4 — FROM CONCEPT TO INDICATOR

The First Discipline of Measurement

- Instrument development starts with conceptual clarity (e.g., firm engagement must be clearly distinguished from customer engagement and employee engagement).
- A construct becomes measurable only when its dimensions are articulated (e.g., masstige marketing had to be decomposed into elements such as prestige, self-congruence, and willingness to pay a premium).
- Indicators should emerge from theory and literature, not guesswork (e.g., digital orientation requires items grounded in the concept of a firm-wide digital configuration).
- The sharper the concept, the sharper the instrument (e.g., vague constructs generate vague items and unstable findings).

Penjelasan

Kesalahan paling umum dalam pengembangan instrumen adalah memulai dari item, bukan dari konsep. Padahal, instrumen yang baik selalu dimulai dari kejelasan konseptual. Peneliti harus bisa menjawab terlebih dahulu apa yang sebenarnya dimaksud oleh konstruk yang ingin diukur, apa batasnya, dan bagaimana ia berbeda dari konstruk lain yang mirip. Artikel *firm engagement: scale development and verification* sangat bagus untuk menjelaskan hal ini, karena ia menegaskan bahwa *firm engagement* tidak boleh disamakan dengan *employee engagement* atau *customer engagement*, walaupun ketiganya tampak berada dalam keluarga konsep yang sama.

Begitu pula dengan *masstige marketing*. Jika konsep ini hanya dibiarkan sebagai istilah populer tentang “prestige for the masses,” pengukurannya akan tetap kabur. Karena itu, penelitian *scale development* di *Journal of Business Research* menguraikannya ke dalam lima dimensi seperti *self-congruence*, *prestige*, *brand knowledge*, *brand loyalty*, dan *willingness to pay a premium*. Ini menunjukkan bahwa pengukuran baru mungkin terjadi ketika peneliti berhasil memecah konsep ke dalam bagian-bagian yang logis dan dapat diamati.

Prinsip yang sama tampak pada *digital orientation*. Artikel *expanding the boundaries of digital orientation research* menekankan bahwa pengukuran lama terlalu bergantung pada proxy berbasis data sekunder, sehingga belum cukup menangkap *digital orientation* sebagai konfigurasi menyeluruh di tingkat organisasi. Karena itu, peneliti mengembangkan survey scale baru setelah terlebih dahulu memperjelas makna konseptualnya. Ini pelajaran yang sangat penting: indikator tidak boleh lahir dari intuisi, tetapi harus tumbuh dari teori.

Maka, hubungan antara konsep dan instrumen bersifat langsung. Jika konsepnya kabur, instrumennya akan kabur. Jika konsepnya tajam, item-itemnya akan lebih terarah, dan hasil penelitiannya akan lebih stabil. Dalam metodologi, kekuatan statistik hampir selalu berawal dari kekuatan konseptual.

5 — CONTENT VALIDITY

Do the Items Adequately Represent the Construct?

- Content validity asks whether the instrument covers the substance of the construct (e.g., entrepreneurial hustle should not be reduced to generic busyness or long working hours).
- It depends on theory, literature review, and expert judgment (e.g., item pools are screened to ensure they reflect the intended conceptual domain).
- Good content validity begins before any statistical test (e.g., weak items cannot be fully repaired later by impressive factor loadings).
- It is especially crucial for new or emerging constructs (e.g., firm engagement and digital orientation require careful conceptual coverage because their boundaries are still evolving).

Penjelasan

Content validity adalah pertanyaan paling mendasar dalam pengukuran: apakah item-item dalam instrumen benar-benar mewakili isi konstruk yang hendak diukur. Ini bukan persoalan teknis kecil, melainkan masalah substansi. Jika konstruk yang diukur adalah *entrepreneurial hustle*, misalnya, lalu item-itemnya hanya bertanya tentang “sibuk bekerja,” maka instrumen itu gagal sejak awal. Artikel *entrepreneurial hustle* justru menekankan bahwa *hustle* berhubungan dengan tindakan yang mendesak dan tidak ortodoks dalam menghadapi peluang dan hambatan di bawah ketidakpastian, bukan sekadar banyaknya aktivitas.

Karena itu, content validity dibangun melalui teori, telaah literatur, dan evaluasi substantif atas item-item yang diajukan. Peneliti harus bertanya: apakah domain konstruk sudah tercakup? Apakah ada aspek penting yang terlewat? Apakah item-item tertentu justru masuk ke wilayah konstruk lain? Artikel firm engagement menunjukkan pentingnya proses ini, karena konstruk tersebut berada dekat dengan customer engagement dan employee engagement sehingga batasnya harus dijaga dengan hati-hati.

Hal yang sangat penting untuk ditekankan kepada pembaca buku adalah bahwa content validity terjadi sebelum statistik. Item yang salah secara konseptual tidak akan sungguh-sungguh menjadi baik hanya karena angka-angka statistiknya tampak rapi. Statistik dapat membantu membersihkan instrumen, tetapi tidak dapat menggantikan kelemahan konseptual yang dibiarkan sejak awal. Inilah sebabnya textbook metodologi selalu menempatkan conceptual definition sebelum measurement analysis.

Content validity menjadi semakin penting ketika konstruk masih baru atau sedang berkembang. Pada situasi seperti itu, peneliti tidak bisa sekadar menyalin item lama, karena batas makna konstruk itu sendiri masih diperdebatkan. Maka, validitas isi adalah garis pertahanan pertama agar instrumen tidak kehilangan makna.

6 — CONSTRUCT VALIDITY

Does the Instrument Really Measure the Intended Construct?

- Construct validity asks whether the instrument truly measures the target concept (e.g., social identity of entrepreneurs should capture self-definition, not just confidence or ambition).
- It usually involves dimensional structure and empirical coherence (e.g., the proposed dimensions should be supported by the data rather than remain theoretical assumptions).
- Construct validity includes convergent and discriminant logic (e.g., related constructs may correlate, but distinct constructs must still remain distinguishable).
- Without construct validity, interpretation becomes unstable (e.g., a high score may not mean what the researcher thinks it means).

Penjelasan

Jika content validity menanyakan apakah isi konstruk sudah tercakup, maka construct validity menanyakan sesuatu yang lebih tajam: apakah instrumen itu benar-benar mengukur konstruk yang dimaksud, bukan sesuatu yang mirip tetapi berbeda. Ini sangat penting karena banyak konstruk dalam riset bisnis tampak berdekatan secara bahasa, tetapi berbeda secara teoretis. Artikel measuring the social identity of entrepreneurs adalah contoh yang sangat baik, karena ia menunjukkan bahwa identitas sosial founder harus dibedakan dari konsep lain seperti self-confidence, passion, atau ambition. Kalau tidak, peneliti akan mengira sedang mengukur identitas, padahal sesungguhnya mengukur hal lain.

Construct validity biasanya diuji melalui struktur faktor dan koherensi dimensional. Tetapi inti logikanya sederhana: bila peneliti mengusulkan bahwa suatu konstruk memiliki dimensi tertentu, maka data harus mendukung bahwa dimensi-dimensi itu memang ada dan memang membentuk konstruk secara masuk akal. Pada artikel masstige marketing, misalnya, lima dimensi yang diusulkan tidak berhenti pada konseptualisasi, tetapi diuji lintas konteks industri seperti restoran, maskapai, dan otomotif. Ini memperlihatkan bahwa struktur konstruk harus dibuktikan, bukan hanya diumumkan.

Construct validity juga bergantung pada logika convergent dan discriminant validity. Konstruk yang berkaitan secara teoritis boleh berhubungan, tetapi tetap harus bisa dibedakan. Artikel work-based learning di Journal of Vocational Behavior memberi ilustrasi penting karena ia melaporkan factorial

structure, measurement invariance, dan dukungan untuk convergent, divergent, and construct validity. Bagi peneliti, ini menunjukkan bahwa validitas konstruk bukan sekadar satu angka, tetapi keseluruhan argumen bahwa alat ukur memang bekerja sebagaimana mestinya.

Tanpa construct validity, interpretasi skor menjadi rapuh. Peneliti mungkin menemukan hubungan yang secara statistik menarik, tetapi tidak pernah benar-benar yakin apa arti skor itu. Karena itu, validitas konstruk adalah syarat agar hasil penelitian dapat dibaca sebagai representasi yang sah dari fenomena yang diteliti.

7 — RELIABILITY

Can the Instrument Produce Consistent Results?

- Reliability concerns the consistency of measurement (e.g., items measuring work-life fulfilment should behave coherently as one scale system).
- Internal consistency is important, but it is not the whole story (e.g., a scale may hang together internally but still shift when the construct itself is unstable over time).
- A reliable instrument is not automatically valid (e.g., consistent items can still measure the wrong thing).
- Reliability gives confidence in the score, but validity gives meaning to the score (e.g., reliability makes measurement dependable; validity makes it interpretable).

Penjelasan

Reliabilitas berbicara tentang sejauh mana instrumen menghasilkan pengukuran yang konsisten dan dapat dipercaya. Dalam arti sederhana, reliabilitas menanyakan apakah item-item dalam suatu instrumen bekerja secara teratur, tidak saling bertabrakan, dan cukup stabil untuk membentuk skor yang masuk akal. Artikel *towards work life fulfilment: scale development and validation* menunjukkan hal ini dengan baik, karena ia melaporkan skala multidimensi yang robust dengan reliabilitas dan validitas yang memadai melalui beberapa studi.

Namun, peneliti perlu berhati-hati untuk tidak menyederhanakan reliabilitas hanya menjadi satu angka seperti Cronbach's alpha. Konsistensi internal memang penting, tetapi ada kasus ketika konstruk yang diukur memang bersifat berubah. Artikel *entrepreneurial fear of failure* menunjukkan bahwa fear of failure tidak stabil sepanjang waktu dan dimensinya berdampak berbeda pada outcome psikologis dan kecenderungan perilaku. Ini memberi pelajaran penting: perubahan skor tidak selalu berarti instrumennya buruk; kadang memang realitas konstruknya bersifat temporal.

Yang lebih penting lagi, reliabilitas tidak identik dengan validitas. Item bisa sangat konsisten satu sama lain, tetapi tetap salah sasaran. Misalnya, skala bisa secara konsisten mengukur antusiasme, padahal peneliti mengira sedang mengukur identity atau engagement. Karena itu, reliabilitas hanya menjawab apakah alat ukur cukup stabil, bukan apakah ia bermakna dengan benar.

Maka, dalam metodologi penelitian, reliabilitas sebaiknya dipahami sebagai syarat kepercayaan terhadap kestabilan pengukuran, sedangkan validitas adalah syarat kepercayaan terhadap makna pengukuran. Penelitian yang kuat membutuhkan keduanya sekaligus.

8 — PILOT TESTING AND SCALE PURIFICATION

Why Good Instruments Are Refined, Not Rushed

- Pilot testing helps detect ambiguity, confusion, and redundancy (e.g., respondents may interpret “firm engagement” very differently if the item wording is vague).
- Scale purification removes weak, overlapping, or misleading items (e.g., items that blur prestige and loyalty should not remain in a prestige scale).
- Refinement improves both validity and reliability (e.g., the final instrument becomes cleaner, sharper, and easier to interpret).
- A polished scale usually emerges through multiple rounds, not one draft (e.g., item generation, expert review, testing, purification, and validation).

Penjelasan

Salah satu kesalahan paling umum dalam penelitian adalah menganggap item pertama yang ditulis sudah cukup layak dipakai. Padahal, pilot testing diperlukan justru karena peneliti sering kali terlalu dekat dengan konstruksinya sendiri. Istilah yang tampak jelas bagi peneliti belum tentu jelas bagi responden. Sebuah item tentang “firm engagement” atau “strategic digital orientation” bisa dibaca dengan arti yang berbeda-beda jika redaksinya terlalu abstrak. Maka, pilot testing berfungsi untuk menguji apakah bahasa instrumen benar-benar dapat dipahami sebagaimana dimaksud.

Setelah itu, peneliti perlu melakukan scale purification. Artinya, item-item yang lemah, tumpang tindih, terlalu umum, atau terlalu dekat dengan konstruk lain harus direvisi atau dibuang. Artikel customer experience in AI-enabled products—yang meskipun tidak kita jadikan contoh utama di sini, sangat jelas menggambarkan proses scale purification, validation, dan cross-testing—menunjukkan bahwa skala yang baik lahir dari pemurnian yang bertahap, bukan dari spontanitas penulisan item. Artikel prestige marketing juga secara eksplisit mengikuti framework scale development ala Churchill, yang menekankan pentingnya refinement sebelum generalisasi.

Pemurnian skala penting karena ia tidak hanya memperbaiki statistik, tetapi juga menjernihkan makna. Ketika item-item yang kabur disingkirkan, konstruk menjadi lebih bersih dan lebih mudah diinterpretasikan. Inilah sebabnya instrument development tidak boleh dipercepat hanya karena peneliti ingin segera mengumpulkan data. Dalam banyak kasus, kualitas instrumen justru menentukan kualitas seluruh penelitian.

Untuk pembaca buku, bagian ini penting karena menunjukkan bahwa instrumen akademik yang baik hampir selalu lahir melalui beberapa putaran pengembangan: mulai dari item generation, expert input, pilot testing, item reduction, hingga validation. Jadi, alat ukur yang tampak sederhana di artikel jurnal sebenarnya adalah hasil kerja metodologis yang sangat serius.

9 — ENTREPRENEURSHIP EXAMPLE

Measuring the Social Identity of Entrepreneurs

- Entrepreneurial identity is not a vague personality trait (e.g., founders differ in how central entrepreneurial roles are to their self-definition).
- A scale can deepen theory by clarifying what is being measured (e.g., social identity helps explain why founders pursue different venture paths).

- International validation strengthens confidence in the construct (e.g., testing across countries reduces the risk that the scale reflects only one local context).
- Good measurement makes entrepreneurial heterogeneity visible (e.g., founders are not all driven by the same self-concept).

Penjelasan

Contoh yang sangat kuat dari bidang entrepreneurship adalah artikel measuring the social identity of entrepreneurs: scale development and international validation di Journal of Business Venturing. Artikel ini penting karena menunjukkan bahwa entrepreneur tidak bisa dipahami secara sederhana sebagai kelompok yang homogen. Perbedaan founder bukan hanya soal sumber daya atau strategi, tetapi juga soal bagaimana mereka mendefinisikan diri mereka sendiri. Maka, pengembangan skala identitas sosial menjadi penting bukan hanya untuk pengukuran, tetapi juga untuk memperjelas teori tentang heterogenitas entrepreneur.

Yang membuat artikel ini menarik adalah argumen dasarnya: tanpa alat ukur yang valid, social identity theory sulit dipakai untuk menjelaskan proses penciptaan venture dan perilaku founder secara lebih sistematis. Dengan kata lain, skala tidak hanya “mengukur sesuatu yang sudah ada,” tetapi membantu membuat suatu konsep menjadi operasional dan produktif secara teoretis. Ini adalah pelajaran penting untuk buku: pengembangan instrumen yang baik dapat memperluas kemampuan teori untuk menjelaskan dunia nyata.

Penelitian ini juga melakukan validasi lintas negara dan wilayah. Langkah seperti ini penting karena ia mengurangi kemungkinan bahwa instrumen hanya mencerminkan konteks lokal tertentu. Dalam metodologi, itu berarti peneliti tidak hanya bertanya “apakah skala ini bekerja,” tetapi juga “apakah skala ini cukup kokoh untuk bertahan melampaui satu konteks budaya.” Di sinilah validitas memperoleh bobot yang lebih besar daripada sekadar performa internal.

Bagi pembaca buku, contoh ini sangat berharga karena menunjukkan bahwa measurement bisa membuat heterogenitas entrepreneur menjadi tampak. Tanpa instrumen yang tepat, perbedaan-perbedaan penting di antara founder akan tenggelam dalam kategori yang terlalu umum.

10 — MARKETING EXAMPLE

Points (English + Example + Indonesian Translation)

- Popular concepts often need sharper measurement, not more repetition (e.g., masstige marketing is widely discussed, but cannot remain only a fashionable label).
- A strong scale forces a concept to become analytically precise (e.g., prestige, self-congruence, and willingness to pay premium must be distinguished rather than blended together).
- Validity matters because managers may act on the results (e.g., a brand strategy based on a flawed scale may misread consumers’ masstige perceptions).
- Generalizability is stronger when the instrument is tested across industries (e.g., restaurant, airline, and automobile contexts reveal whether the scale travels well).

Penjelasan

Contoh yang sangat menarik dari bidang marketing adalah artikel *Masstige Marketing: A scale development and validity study* di *Journal of Business Research*. *Masstige* adalah konsep yang sangat populer dalam praktik, tetapi popularitas tidak otomatis berarti kejernihan konseptual. Justru karena sering digunakan, konsep ini berisiko menjadi kabur. Di sinilah *scale development* berperan: bukan hanya untuk “mengukur *masstige*,” tetapi untuk memaksa konsep itu menjadi analitis, presisi, dan terstruktur.

Artikel ini mengembangkan *MasCal* sebagai skala lima dimensi— *self-congruence*, *prestige*, *brand knowledge*, *brand loyalty*, dan *willingness to pay a premium*. Pelajaran metodologisnya sangat penting: sebuah konsep yang tampak sederhana di permukaan ternyata perlu dipecah menjadi komponen yang berbeda agar dapat diukur dengan benar. Jika semua unsur itu dicampur, peneliti mungkin mendapatkan skor, tetapi tidak benar-benar mendapatkan pemahaman.

Kekuatan lain dari penelitian ini adalah pengujiannya di beberapa industri, yaitu restoran, maskapai, dan otomotif. Ini bukan detail kecil. Pengujian lintas konteks membantu menunjukkan bahwa instrumen tidak hanya cocok untuk satu setting sempit. Bagi metodologi, ini berarti validitas dan generalizability dibangun secara lebih serius, bukan hanya diasumsikan dari satu sampel tunggal.

Bagi dunia praktik, contoh ini juga mengingatkan bahwa skala yang lemah bisa menyesatkan keputusan manajerial. Bila manajer salah membaca apa yang dianggap “*masstige*” oleh konsumen, strategi positioning dapat meleset. Karena itu, pengembangan instrumen bukan hanya urusan akademik, tetapi juga berkaitan dengan kualitas keputusan bisnis.

11 — DIGITAL / SERVICE / HRM EXAMPLE

Measuring Frontline Employee Empowerment in Analytics-Driven Services

- Old constructs may need new instruments in new settings (e.g., empowerment in analytics-driven services is not identical to empowerment in traditional service environments).
- Scale development must follow contextual change (e.g., digital tools, information access, and analytics reshape what empowerment means for frontline employees).
- Mixed evidence-building can strengthen the scale (e.g., literature review, interviews, and survey-based validation may work together).
- Measurement quality matters because service performance may depend on it (e.g., mismeasuring empowerment may lead firms to redesign jobs in the wrong direction).

Penjelasan

Contoh yang sangat relevan dari bidang service dan HRM adalah artikel *frontline employee empowerment: scale development and validation using Confirmatory Composite Analysis* di *International Journal of Information Management*. Artikel ini penting karena mengingatkan kita bahwa konstruk yang sudah lama dikenal dalam teori tidak selalu bisa diukur dengan cara lama dalam konteks baru. Empowerment pada layanan berbasis analitik jelas tidak identik dengan empowerment pada layanan tradisional, karena kini ada dimensi *tools*, *technologies*, and *information access* yang menjadi jauh lebih sentral.

Penelitian ini juga menarik secara metodologis karena tidak langsung melompat ke survei. Peneliti lebih dulu melakukan *systematic literature search* untuk mengidentifikasi dimensi empowerment, lalu melengkapinya dengan *semi-structured interviews* sebelum masuk ke *survey-based validation*. Ini contoh

yang sangat baik untuk menunjukkan bahwa pengembangan instrumen yang kuat sering kali menggabungkan kerja konseptual, kerja kualitatif, dan kerja kuantitatif secara berurutan.

Yang paling penting, artikel ini menunjukkan bahwa perubahan konteks dapat mengubah makna konstruk. Ketika layanan menjadi analytics-driven, empowerment tidak lagi hanya berarti “otonomi membuat keputusan,” tetapi juga berkaitan dengan knowledge, information access, training, dan tools yang memungkinkan frontline employees bertindak secara efektif. Jadi, instrument development bukan hanya pekerjaan mengukur ulang, tetapi juga pekerjaan memikirkan ulang makna konstruk dalam kondisi baru.

Bagi penelitian dan praktik, ini sangat penting. Jika perusahaan mengukur empowerment dengan alat yang tidak lagi sesuai konteks, maka kebijakan SDM dan desain pekerjaan bisa bergerak ke arah yang salah. Artinya, kualitas pengukuran punya konsekuensi nyata, bukan sekadar statistik yang rapi di artikel ilmiah.

12 — COMMON MISTAKES IN INSTRUMENT DEVELOPMENT

Why Many Studies Look Strong but Measure Weakly

- Many researchers jump too quickly from variable names to questionnaire items (e.g., writing “digital mindset” items before defining what the construct actually means).
- Popularity of a construct does not guarantee quality of its measurement (e.g., engagement, resilience, and empowerment can all be overused but poorly operationalized).
- Researchers often overtrust statistics and underinvest in conceptual clarity (e.g., a high loading does not automatically prove that an item is substantively correct).
- Weak measurement undermines the credibility of the whole model (e.g., elegant hypotheses collapse when the constructs are poorly captured).

Penjelasan

Salah satu kesalahan paling sering dalam penelitian adalah memulai dari nama variabel yang terdengar menarik, lalu segera menulis item kuesioner tanpa membangun definisi konseptual yang cukup. Ini membuat instrumen tampak lengkap di permukaan, tetapi kosong secara substantif. Artikel-artikel scale development di jurnal Scopus bereputasi justru memperlihatkan kebalikannya: peneliti menghabiskan energi besar untuk konseptualisasi, dimensionalisasi, dan pemurnian item sebelum data utama dikumpulkan.

Kesalahan lain adalah mengira bahwa karena suatu konsep populer, maka pengukurannya sudah otomatis baik. Ini tidak benar. Engagement, empowerment, social identity, digital orientation, dan masstige semua tampak akrab dalam literatur, tetapi masing-masing tetap memerlukan kerja pengukuran yang serius. Popularitas tidak pernah menggantikan ketepatan konseptual. Justru konsep yang paling populer sering paling rentan diukur secara longgar.

Banyak peneliti juga terlalu cepat percaya pada statistik. Padahal angka yang baik tidak selalu berarti makna yang baik. Factor loading yang tinggi, misalnya, tidak otomatis membuktikan bahwa item itu benar secara konseptual. Ia hanya menunjukkan bahwa item tersebut cocok secara pola dengan item lain. Jika isi item dari awal sudah salah arah, statistik yang rapi hanya akan memberi ilusi kualitas.

Pada akhirnya, pengukuran yang lemah akan menarik turun seluruh model penelitian. Hipotesis yang canggih, analisis yang kompleks, bahkan sampel yang besar tidak akan mampu menyelamatkan

penelitian yang tidak benar-benar tahu apa yang sedang diukur. Karena itu, instrument development harus diperlakukan sebagai jantung penelitian, bukan lampiran metodologis.

13 — CONCLUSION

- Instrument development is where conceptual rigor becomes measurement rigor (e.g., abstract constructs must be turned into defensible empirical indicators).
- Validity asks whether the instrument means what it claims to mean (e.g., whether a scale really captures identity, engagement, or empowerment rather than something adjacent).
empowerment, bukan sesuatu yang hanya mirip).
- Reliability asks whether the instrument is consistent enough to be trusted (e.g., whether the score can be treated as dependable rather than erratic).
Terjemahan: Reliabilitas menanyakan apakah instrumen cukup konsisten untuk dipercaya (misalnya, apakah skor dapat diperlakukan sebagai sesuatu yang dapat diandalkan, bukan serba acak).
- Strong research requires instruments that are both theoretically grounded and empirically defensible (e.g., a weak instrument can damage even a strong model).

Penjelasan

Pada akhirnya, instrument development adalah proses yang membuat ketelitian konseptual berubah menjadi ketelitian pengukuran. Peneliti tidak cukup hanya memiliki teori yang bagus; ia juga harus menunjukkan bahwa teori itu dapat dihadirkan dalam bentuk indikator yang benar-benar masuk akal, terstruktur, dan dapat diuji. Di titik inilah pengukuran menjadi bagian dari integritas ilmiah.

Validitas dan reliabilitas lalu menjadi dua pertanyaan besar yang tidak boleh dipisahkan. Validitas bertanya apakah instrumen benar-benar berarti seperti yang kita pikirkan. Reliabilitas bertanya apakah instrumen itu cukup konsisten untuk dipercaya. Jika salah satu lemah, maka kualitas inferensi ikut melemah. Penelitian yang kuat membutuhkan alat ukur yang tidak hanya stabil, tetapi juga benar.

Kita juga telah melihat bahwa instrument development bisa menjadi kontribusi ilmiah tersendiri. Dengan memperjelas cara suatu konsep diukur, peneliti sering kali sekaligus memperjelas bagaimana konsep itu dipahami. Jadi, scale development bukan hanya urusan metode; ia juga sering menjadi kerja teoritis.

■ PART 14 — KEY TAKEAWAYS & CLOSING REMARKS

- A good instrument begins with a good concept (e.g., unclear constructs produce unclear items).
- Validity must be built, not assumed (e.g., item relevance, dimensional structure, and construct boundaries all matter).
- Reliability supports confidence, but does not replace meaning (e.g., a consistent scale can still be wrong in substance).
- The next step is to learn how trustworthy instruments support quantitative analysis (e.g., moving from good measurement to sound statistical interpretation).

Penjelasan

Pertemuan ini membawa kita pada satu pelajaran utama: kualitas penelitian tidak hanya bergantung pada pertanyaan yang bagus atau teori yang menarik, tetapi juga pada alat ukur yang dipakai untuk menangkap fenomena. Peneliti yang matang akan selalu sadar bahwa measurement bukan pekerjaan mekanis, melainkan bagian dari logika ilmiah itu sendiri.

Karena itu, peneliti harus belajar memperlakukan instrumen dengan serius. Ia harus dibangun dari konsep yang jelas, diuji dengan hati-hati, dimurnikan secara bertahap, dan dibaca kembali dengan reflektif. Tanpa disiplin ini, penelitian mudah terlihat meyakinkan di permukaan tetapi rapuh di fondasi.

See you in Meeting 10, with the topic:

Quantitative Data Analysis: Descriptive Statistics and Interpretation

Pada pertemuan berikutnya, kita akan bergerak dari how constructs are measured menuju how the resulting data are read, summarized, and interpreted, sehingga peneliti tidak hanya memiliki skor, tetapi juga mampu membaca makna statistiknya secara benar.

reference

1. Research Methods for Business Students, Mark N. K. Saunders, Philip Lewis, Adrian Thornhill, Pearson, 2024, 888 pages.
2. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, John W. Creswell, J. David Creswell, SAGE Publications, 2018, 304 pages.
3. Research Methods For Business: A Skill Building Approach, Roger Bougie, Uma Sekaran, John Wiley & Sons, 2020, 432 pages.
4. Business Research Methods, Emma Bell, Bill Harley, Alan Bryman, Oxford University Press, 2022, 647 pages.
5. Measuring the Social Identity of Entrepreneurs: Scale Development and International Validation, Melissa S. Cardon, Joakim Wincent, Jagdip Singh, Mateja Drnovšek, Journal of Business Venturing, Elsevier, 2016, 542-572.
6. Entrepreneurial Fear of Failure: Scale Development and Validation, Gabriella Cacciotti, James C. Hayton, J. Robert Mitchell, David G. Allen, Journal of Business Venturing, Elsevier, 2020, 106041.
7. Entrepreneurial Hustle: Scale Development and Validation, Devin Burnell, Emily Neubert, Greg Fisher, Matthew R. Marvel, Regan Stevenson, Donald F. Kuratko, Journal of Business Venturing, Elsevier, 2024, 106407.
8. Masstige Marketing: A Scale Development and Validity Study, Muhammad Ishtiaq Ishaq, Ali Raza, Boris Bartikowski, Huma Sarwar, Journal of Business Research, Elsevier, 2023, 114112.
9. Frontline Employee Empowerment: Scale Development and Validation Using Confirmatory Composite Analysis, Saradhi Motamarri, Shahriar Akter, Venkata Yanamandram, International Journal of Information Management, Elsevier, 2020, 102177.
10. Firm Engagement: Scale Development and Verification, Xiaoyun Han, Yundan Zheng, Catherine Prentice, Shuping Chen, Journal of Retailing and Consumer Services, Elsevier, 2024, 103902.
11. Work-Based Learning: Development and Validation of a Scale Measuring the Learning Potential of the Workplace (LPW), Irina Nikolova, Joris van Ruysseveldt, Hans De Witte, Jef Syroit, Journal of Vocational Behavior, Elsevier, 2014, 1-10.
12. Towards Work Life Fulfilment: Scale Development and Validation, Puja Khatri, Shalu Shukla, Asha Thomas, Atul Shiva, Abhishek Behl, Journal of Business Research, Elsevier, 2025, 115006.
13. Expanding the Boundaries of Digital Orientation Research: Scale Development and Validation, Bastian Kindermann, Corinna Vera Hedwig Schmidt, Florian Fengel, Steffen Strese, Journal of Business Research, Elsevier, 2024, 114895.