

Pengembangan dan Validasi *Fraud Risk Profiling Scale*: Pendekatan Situational Judgment Test Berbasis *Fraud Diamond Theory****Development and Validation of the Fraud Risk Profiling Scale: A Situational Judgment Test Approach Based on Fraud Diamond Theory*****Fathul Lubabin Nuqul¹, Fidinda Avitasari²,**

1 Universitas Islam Negeri Maulana Malik Ibrahim Malang

2 Universitas Islam Negeri Maulana Malik Ibrahim Malang

e-mail: lubabin_nuqul@uin-malang.ac.id

ABSTRAK

Kecurangan dalam keuangan (*fraud*) merupakan ancaman serius terhadap integritas organisasi di berbagai sektor, baik publik maupun swasta. Meskipun kerangka teoretis *Fraud Diamond*, yang mencakup tekanan (*pressure*), rasionalisasi (*rationalization*), kesempatan (*opportunity*), dan kapabilitas (*capability*), telah diakui secara luas sebagai model komprehensif dalam memahami risiko *fraud*, namun instrumen pengukuran yang mengintegrasikan keempat dimensi tersebut masih terbatas. Penelitian ini bertujuan mengembangkan dan menguji validitas psikometrik instrumen profil risiko *fraud* berbasis *Fraud Diamond* dengan pendekatan Situational Judgment Test (SJT). Format SJT dipilih untuk meminimalkan bias keinginan sosial (*social desirability bias*) dan meningkatkan realisme respons perilaku. Dari 40 item awal, melalui validasi ahli dan uji coba awal, disaring menjadi 14 item yang diujikan kepada 252 mahasiswa aktif organisasi di salah satu perguruan tinggi di Indonesia. Analisis Faktor Konfirmatori (CFA) menggunakan AMOS menunjukkan kesesuaian model yang sangat baik ($\chi^2/df = 1.67$, GFI = 0.952, AGFI = 0.929, RMSEA = 0.038, RMR = 0.028), serta reliabilitas yang memadai ($\alpha = 0.803$). Meskipun nilai loading factor berkisar antara 0.427–0.656. Hal ini menunjukkan potensi dan menyisakan tantangan untuk penyempurnaan. Temuan ini mendukung validitas konstruk instrumen sebagai alat skrining awal risiko *fraud*. Penelitian ini memberikan kontribusi penting dalam pengembangan alat ukur berbasis bukti yang dapat digunakan dalam konteks pendidikan, organisasi, maupun pencegahan korupsi. Rekomendasi untuk penelitian lanjutan mencakup validasi silang pada populasi yang lebih beragam dan konteks budaya yang berbeda guna memperkuat generalisasi temuan.

Kata kunci :Risiko *Fraud*, *Fraud Diamond*, Situational Judgment Test, Pengembangan Instrumen, Penilaian Integritas, Etika Perilaku**ABSTRACT**

Financial fraud poses a serious threat to the integrity of organizations across both public and private sectors. Although the *Fraud Diamond* theoretical framework, encompassing pressure, rationalization, opportunity, and capability, has been widely recognized as a comprehensive model for understanding fraud risk, measurement instruments integrating these four dimensions remain limited. This study aims to develop and test the psychometric validity of a fraud risk profile instrument based on the *Fraud Diamond* using the Situational Judgment Test (SJT) approach. The SJT format was chosen to minimize social desirability bias and increase the realistic nature of behavioral responses. From the initial 40 items, through expert validation and initial pilot testing, 14 items were refined and tested on 252 active students at a university in Indonesia. Confirmatory Factor Analysis (CFA) using AMOS showed excellent model fit ($\chi^2/df = 1.67$, GFI = 0.952, AGFI = 0.929, RMSEA = 0.038, RMR = 0.028), and adequate reliability ($\alpha = 0.803$). Although the factor loading values ranged from 0.427–0.656, this indicates potential and leaves challenges for improvement. These findings support the construct validity of the instrument as an initial fraud risk screening tool. This study provides an important contribution to the development of evidence-based measurement tools that can be used in educational, organizational, and corruption prevention contexts. Recommendations for further research include cross-validation in more diverse populations and different cultural contexts to strengthen the generalizability of the findings.

Keywords :Fraud Risk, *Fraud Diamond*, Situational Judgment Test, Instrument Development, Integrity Assessment, Behavioral Ethics

Copyright © 2025 JPFI: Jurnal Psikologi Forensik Indonesia



This is an open article under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

PENDAHULUAN

Fraud atau kecurangan didefinisikan sebagai tindakan penipuan yang disengaja untuk memperoleh keuntungan yang tidak sah atau merugikan pihak lain (*Association of Certified Fraud Examiners* (ACFE, 2022). Fenomena *fraud* telah menjadi ancaman serius bagi integritas organisasi, institusi pendidikan, dan bahkan negara. Di Indonesia, data Komisi Pemberantasan Korupsi (KPK) menunjukkan bahwa Indeks Persepsi Korupsi Indonesia mengalami penurunan dari 71.94 pada tahun 2022 menjadi 70.97 pada tahun 2023, mengindikasikan peningkatan kerentanan terhadap praktik korupsi dan *fraud* (KPK, 2023). Lebih mengkhawatirkan lagi, berbagai kasus *fraud* di sektor publik dan swasta menunjukkan bahwa perilaku ini tidak hanya terjadi pada level profesional, tetapi juga berakar pada perilaku yang terbentuk sejak masa pendidikan.

Dalam konteks pendidikan tinggi, *academic fraud* atau kecurangan akademik telah menjadi fenomena yang mengkhawatirkan. Survei global menunjukkan bahwa 68% mahasiswa sarjana mengakui telah melakukan perilaku tidak jujur dalam akademik pada tahun 2022, meningkat dari 60% pada tahun 2020 (International Center for Academic Integrity, 2022). Di Indonesia, penelitian menunjukkan bahwa 45-60% mahasiswa pernah terlibat dalam berbagai bentuk kecurangan akademik, mulai dari plagiarisme, menyontek, hingga manipulasi data penelitian (Amelia & Setiawan, 2021). Lebih jauh lagi, mahasiswa yang aktif dalam organisasi kemahasiswaan menghadapi risiko tambahan berupa *fraud* organisasional, seperti penyalahgunaan dana organisasi, manipulasi laporan kegiatan, dan nepotisme dalam pengambilan keputusan.

Fenomena ini menunjukkan bahwa perilaku *fraud* tidak muncul secara tiba-tiba pada saat seseorang memasuki dunia kerja, tetapi merupakan hasil dari proses pembelajaran dan habituasi yang dimulai sejak masa pendidikan. Oleh karena itu, identifikasi dini terhadap faktor-faktor risiko *fraud* pada populasi mahasiswa menjadi sangat penting sebagai upaya pencegahan dan intervensi *early warning*. Namun, upaya identifikasi ini memerlukan instrumen pengukuran yang valid, reliabel, dan berbasis pada kerangka teoretis yang kuat.

Pemahaman tentang mengapa seseorang melakukan *fraud* telah menjadi fokus penelitian selama beberapa dekade. Teori yang paling berpengaruh dalam menjelaskan perilaku *fraud* adalah *Fraud Triangle* yang dikemukakan oleh Donald Cressey pada tahun 1953, yang mengidentifikasi tiga elemen kunci: *pressure* (tekanan), *opportunity* (kesempatan), dan *rationalization* (rasionalisasi) (Cressey, 1953). Namun, teori ini kemudian disempurnakan oleh Wolfe dan Hermanson (2004) menjadi *Fraud Diamond Theory* dengan menambahkan elemen keempat, yaitu *capability* (kapabilitas), yang mengakui bahwa tidak semua individu yang menghadapi tekanan, kesempatan, dan rasionalisasi akan melakukan *fraud* diperlukan kemampuan atau kapabilitas tertentu untuk mengeksekusi tindakan *fraud*.

Pressure (Tekanan) merujuk pada motivasi atau dorongan yang dirasakan individu untuk melakukan *fraud*. Tekanan dapat berasal dari faktor internal maupun eksternal. Tekanan internal mencakup kondisi finansial pribadi yang mendesak, kebutuhan untuk mempertahankan gaya hidup tertentu, atau tekanan psikologis seperti kecemasan dan stres (Reurink, 2018). Tekanan eksternal dapat berupa target kinerja yang tidak realistis, tuntutan dari atasan atau rekan, atau ekspektasi sosial yang tinggi (Free, 2015). Dalam konteks mahasiswa, tekanan dapat berupa kebutuhan untuk mempertahankan prestasi akademik, tuntutan untuk sukses dalam organisasi, atau tekanan finansial untuk membiayai kegiatan (Dias-Oliveira et al., 2024).

Opportunity (Kesempatan) mengacu pada situasi atau kondisi yang memungkinkan *fraud* terjadi tanpa terdeteksi. Kesempatan muncul ketika sistem kontrol internal lemah, pengawasan tidak memadai, atau terdapat celah dalam prosedur yang dapat dieksploitasi (Lokanan, 2015). Dalam organisasi mahasiswa, kesempatan dapat berupa akses tidak terbatas terhadap dana organisasi, prosedur pencairan dana yang longgar, atau tidak adanya mekanisme audit internal yang efektif (Rahman et al., 2025). Penelitian menunjukkan bahwa kelemahan dalam sistem pengawasan merupakan prediktor signifikan terhadap terjadinya *fraud* (Albrecht et al., 2019).

Rationalization (Rasionalisasi) adalah proses kognitif di mana individu membenarkan tindakan *fraud* mereka sehingga tindakan tersebut tampak

dapat diterima secara moral. Rasionalisasi memungkinkan individu untuk melakukan *fraud* tanpa merasa bersalah atau melanggar nilai-nilai pribadi mereka (Anand et al., 2015). Bentuk rasionalisasi yang umum termasuk "semua orang melakukannya", "ini hanya pinjaman sementara", "organisasi tidak akan rugi", atau "saya berhak mendapatkan ini karena kerja keras saya" (Murphy & Free, 2016). Dalam konteks akademik, mahasiswa sering merasionalisasi kecurangan dengan alasan seperti "sistem penilaian tidak adil", "dosen tidak kompeten", atau "ini hanya untuk lulus, bukan untuk kehidupan nyata" (Dias-Oliveira et al., 2024).

Capability (Kapabilitas) adalah elemen yang ditambahkan oleh Wolfe dan Hermanson (2004) yang mengakui bahwa *fraud* memerlukan keterampilan, pengetahuan, dan posisi tertentu untuk dapat dieksekusi. Kapabilitas mencakup pemahaman tentang sistem dan prosedur, kemampuan untuk mengeksploitasi kelemahan, keterampilan untuk menutupi jejak, dan posisi atau otoritas yang memungkinkan akses terhadap sumber daya (Wolfe & Hermanson, 2004). Dalam konteks mahasiswa, kapabilitas dapat berupa posisi kepemimpinan dalam organisasi (ketua, bendahara), pemahaman tentang sistem administrasi kampus, atau keterampilan teknis seperti kemampuan manipulasi dokumen (Baz et al., 2016).

Fraud Diamond Theory telah banyak digunakan dalam penelitian *fraud* di berbagai konteks, termasuk *fraud* korporat (Rahman et al., 2025), *fraud* akademik (Dias-Oliveira et al., 2024), dan *fraud* sektor publik (Darwis & Samukri, 2024). Namun, mayoritas penelitian menggunakan *Fraud Diamond* sebagai kerangka analisis untuk menjelaskan kasus *fraud* yang sudah terjadi, bukan sebagai dasar untuk mengembangkan instrumen prediktif yang dapat mengidentifikasi risiko *fraud* sebelum terjadi. Penelitian ini berupaya mengisi kesenjangan tersebut dengan mengembangkan instrumen pengukuran risiko *fraud* yang berbasis pada keempat dimensi *Fraud Diamond*. Meskipun *Fraud Diamond Theory* telah banyak digunakan dalam penelitian, instrumen pengukuran yang komprehensif dan tervalidasi dengan baik untuk menilai risiko *fraud* masih terbatas. Tinjauan literatur mengidentifikasi beberapa instrumen yang telah dikembangkan, namun masing-masing memiliki keterbatasan.

Pertama, *Fraud Risk Assessment Questionnaire* yang dikembangkan oleh ACFE (2020)

merupakan instrumen yang paling banyak digunakan dalam praktik audit. Namun, instrumen ini dirancang untuk auditor profesional dalam menilai risiko *fraud* organisasi, bukan untuk mengukur kecenderungan individu terhadap perilaku *fraud*. Selain itu, instrumen ini tidak berbasis pada teori psikologis yang kuat dan lebih bersifat *checklist* prosedural daripada pengukuran konstruk laten.

Kedua, *Academic Misconduct Scale* yang dikembangkan oleh berbagai peneliti (misalnya, Veríssimo et al., 2025; Fritz et al., 2024) fokus pada pengukuran perilaku kecurangan akademik. Instrumen-instrumen ini umumnya menggunakan format *self-report* yang menanyakan frekuensi perilaku kecurangan yang pernah dilakukan (misalnya, "Seberapa sering Anda menyontek dalam ujian?"). Kelemahan utama format ini adalah rentan terhadap *social desirability* bias, di mana responden cenderung *under-report* perilaku yang tidak diinginkan secara sosial (Paulhus & Vazire, 2007). Selain itu, instrumen-instrumen ini umumnya hanya mengukur satu atau dua dimensi dari *Fraud Triangle/Diamond* (biasanya *pressure* dan *opportunity*), tidak mencakup keempat dimensi secara komprehensif.

Ketiga, *Fraud Risk Judgment Scale* yang dikembangkan oleh Julian et al. (2022) dirancang untuk mengukur kemampuan auditor internal dalam menilai risiko *fraud*. Instrumen ini memiliki properti psikometrik yang baik (reliabilitas dan validitas yang memadai) dan berbasis pada teori *fraud*. Namun, target populasinya adalah profesional auditor, bukan mahasiswa atau populasi umum. Selain itu, instrumen ini mengukur kemampuan *judgment* terhadap skenario *fraud*, bukan kecenderungan atau risiko individu untuk melakukan *fraud*.

Keempat, beberapa penelitian telah menggunakan adaptasi dari skala integritas atau etika untuk mengukur risiko *fraud* secara tidak langsung (misalnya, *Integrity Test*, *Counterproductive Work Behavior Scale*). Namun, konstruk integritas dan etika tidak identik dengan risiko *fraud*. Seseorang dapat memiliki integritas tinggi dalam konteks umum tetapi tetap berisiko melakukan *fraud* dalam situasi tertentu yang melibatkan tekanan ekstrem, kesempatan yang sangat besar, dan rasionalisasi yang kuat. Tabel 1 menyajikan perbandingan instrumen-instrumen yang sudah ada dengan instrumen yang dikembangkan dalam penelitian ini.

Tabel 1
Perbandingan Instrumen Pengukuran *Fraud* yang Sudah Ada

Instrumen	Penulis	Dimensi yang Diukur	Jumlah Item	Format	Populasi Target	Reliabilitas	Keterbatasan Utama
Fraud Risk Assessment Questionnaire	ACFE (2020)	Organizational fraud risk factors	50+	Checklist	Auditor profesional	Tidak dilaporkan	Bukan pengukuran konstruk psikologis; tidak berbasis teori
Academic Misconduct Questionnaire	Verissi mo et al. (2025)	Cheating behavior, plagiarism	18	Self-report frequency	Mahasiswa	$\alpha = 0.82$	Rentan social desirability bias; hanya 2 dimensi
Fraud Risk Judgment Scale	Julian et al. (2022)	Fraud risk judgment ability	15	Likert scale	Auditor internal	$\alpha = 0.87$	Mengukur kemampuan judgment, bukan risiko individu
Integrity Test	Ones & Viswesv aran (1998)	General integrity	25-100	True/False	Karyawan	$\alpha = 0.75-0.85$	Konstruk berbeda dengan fraud risk; tidak spesifik

Dari tinjauan ini, jelas bahwa terdapat kebutuhan untuk mengembangkan instrumen yang: (1) berbasis pada kerangka teoretis *Fraud Diamond* yang komprehensif, mencakup keempat dimensi; (2) menggunakan format yang dapat mengurangi *social desirability* bias; (3) dirancang khusus untuk populasi mahasiswa dengan konteks yang relevan; dan (4) memiliki properti psikometrik yang tervalidasi dengan baik. Penelitian ini berupaya memenuhi kebutuhan tersebut.

Untuk mengatasi masalah bias *social desirability* yang umum terjadi pada instrumen *self-report*, penelitian ini menggunakan format *Situational Judgment Test* (SJT). SJT adalah metode pengukuran yang menyajikan skenario situasional yang realistis dan meminta responden untuk mengevaluasi atau memilih respons yang paling mungkin mereka lakukan dalam situasi tersebut (Whetzel et al., 2020). Format SJT memiliki beberapa keunggulan dibandingkan format *self-report* tradisional.

Pertama, SJT dapat mengurangi bias *social desirability* karena responden tidak secara eksplisit ditanya tentang perilaku mereka sendiri, tetapi diminta untuk menilai atau memilih respons terhadap situasi hipotetis (McDaniel et al., 2007). Penelitian menunjukkan bahwa SJT memiliki korelasi yang lebih rendah dengan skala *social desirability* dibandingkan dengan kuesioner *self-report* langsung (Nguyen et al., 2005). Dalam

konteks pengukuran *fraud risk*, ini sangat penting karena responden cenderung tidak mau mengakui kecenderungan mereka untuk melakukan *fraud*.

Kedua, SJT memiliki validitas prediktif yang baik terhadap perilaku aktual. Meta-analisis menunjukkan bahwa SJT memiliki validitas kriteria yang moderat hingga tinggi ($p = 0.26-0.34$) dalam memprediksi kinerja pekerjaan dan perilaku kontraproduktif (Christian et al., 2010). Dalam konteks pendidikan, SJT telah terbukti valid dalam memprediksi profesionalisme dan perilaku etis mahasiswa (Smith et al., 2020; Aylott et al., 2023).

Ketiga, SJT dapat mengukur konstruk yang kompleks dan kontekstual. *Fraud risk* bukan konstruk unidimensional yang sederhana, tetapi melibatkan interaksi antara karakteristik individu, situasi, dan konteks. SJT memungkinkan pengukuran konstruk ini dengan cara yang lebih ekologis valid dibandingkan item *Likert scale* yang abstrak (Lievens & Motowidlo, 2016).

Keempat, SJT memiliki *face validity* yang tinggi. Responden umumnya menganggap SJT sebagai metode pengukuran yang adil dan relevan karena skenario yang disajikan realistis dan terkait dengan konteks mereka (Whetzel et al., 2020). Ini penting untuk meningkatkan motivasi responden dan mengurangi reaktansi.

Namun, pengembangan SJT juga memiliki tantangan. Tantangan utama adalah dalam

penentuan *scoring key*, yaitu bagaimana menentukan respons mana yang "benar" atau "lebih baik" (Whetzel et al., 2020). Terdapat dua pendekatan utama dalam *scoring* SJT: (1) *expert consensus*, di mana panel ahli menentukan respons yang paling tepat berdasarkan pengetahuan profesional mereka; dan (2) *empirical keying*, di mana *scoring* ditentukan berdasarkan respons dari kelompok kriteria (misalnya, individu dengan kinerja tinggi vs rendah) (De Leng et al., 2017). Penelitian menunjukkan bahwa kedua metode memiliki validitas yang sebanding, tetapi *expert consensus* lebih praktis dan dapat diterapkan sebelum pengumpulan data.

Tantangan kedua adalah dalam pengembangan skenario yang realistis, relevan, dan mencakup semua dimensi konstruk yang diukur. Skenario harus cukup spesifik untuk membangkitkan respons yang bermakna, tetapi tidak terlalu spesifik sehingga hanya relevan bagi sebagian kecil responden. Dalam penelitian ini, skenario dikembangkan berdasarkan critical incidents yang diperoleh melalui wawancara dengan mahasiswa aktivis organisasi.

METODE PENELITIAN

Desain Penelitian

Penelitian ini menggunakan desain pengembangan dan validasi instrumen (*scale development and validation study*) dengan pendekatan kuantitatif. Proses pengembangan mengikuti *framework* tiga fase yang direkomendasikan oleh Boateng et al. (2018): (1) *Item Development*, (2) *Scale Development*, dan (3) *Scale Evaluation*. Penelitian ini bersifat *cross-sectional*, di mana data dikumpulkan pada satu titik waktu.

Fase 1: Pengembangan Item

Pada fase pengembangan item diawali dengan menyusun item yang terdiri dari skenario dan jawaban, pengujian validitas konten dan menyusun bobot skor untuk masing-masing jawaban

1.1. Penyusunan Skenario dan Jawaban

Peneliti mengembangkan 40 skenario SJT berdasarkan konteks kehidupan organisasi kemahasiswaan. Setiap skenario menggambarkan situasi dilematis yang realistis dalam konteks kehidupan mahasiswa dan organisasi kemahasiswaan. Struktur setiap skenario terdiri dari: (a) konteks situasi yang menjelaskan latar belakang dan peran responden; (b) dilema atau masalah yang dihadapi; dan (c) empat pilihan respons yang bervariasi dalam tingkat kesesuaian dengan standar etika. Keempat pilihan respons dirancang

untuk merepresentasikan gradasi dari respons yang jelas-jelas merupakan *fraud* (skor 1), respons yang ambigu atau rasionalisasi (skor 2-3), hingga respons yang etis dan sesuai prosedur (skor 4).

Distribusi 40 item awal adalah: *Pressure* (12 item), *Opportunity* (10 item), *Rationalization* (10 item), dan *Capability* (8 item). Distribusi ini mempertimbangkan kompleksitas masing-masing dimensi dan hasil wawancara kualitatif yang menunjukkan bahwa *pressure* dan *opportunity* adalah dimensi yang paling banyak muncul dalam pengalaman mahasiswa.

1.2. Validasi Konten

Validasi konten dilakukan untuk memastikan bahwa item-item yang dikembangkan relevan dan representatif terhadap konstruk yang diukur (Boateng et al., 2018). Proses validasi melibatkan panel ahli yang terdiri dari tiga orang dengan kualifikasi. Setiap ahli diminta untuk menilai setiap item pada dua aspek: (1) *Relevansi*: Seberapa relevan item tersebut dengan dimensi *Fraud Diamond* yang dituju (skala 1-4: 1 = tidak relevan, 2 = agak relevan, 3 = relevan, 4 = sangat relevan); dan (2) *Clarity*: Seberapa jelas dan mudah dipahami skenario dan pilihan respons (skala 1-4: 1 = tidak jelas, 2 = agak jelas, 3 = jelas, 4 = sangat jelas). Ahli juga diminta memberikan *feedback* kualitatif untuk perbaikan redaksional. Dari 40 item awal, 27 item memenuhi kriteria, Tiga belas item yang tidak memenuhi kriteria dieliminasi atau direvisi berdasarkan *feedback* ahli. Alasan utama eliminasi adalah: (a) skenario terlalu spesifik dan tidak relevan bagi sebagian besar mahasiswa (5 item); (b) pilihan respons tidak jelas atau ambigu (4 item); (c) item mengukur lebih dari satu dimensi (*overlap*) (3 item); dan (d) skenario tidak realistis (1 item). S-CVI untuk 27 item yang dipertahankan adalah 0.89, mengindikasikan validitas konten yang sangat baik. Distribusi 27 item setelah validasi konten adalah: *Pressure* (8 item), *Opportunity* (7 item), *Rationalization* (7 item), dan *Capability* (5 item).

1.3. Penentuan Scoring Key

Penentuan *scoring key* merupakan langkah kritis dalam pengembangan SJT (Whetzel et al., 2020). Penelitian ini menggunakan metode *expert consensus* untuk menentukan *scoring*. Proses ini dilakukan *Briefing Expert*. Melibatkan tiga ahli yang sama yang terlibat dalam validasi konten diminta untuk berpartisipasi dalam penentuan *scoring*. Peneliti memberikan *briefing* tentang tujuan instrumen, definisi operasional masing-masing dimensi *Fraud Diamond*, dan kriteria untuk menilai kesesuaian respons.

Fase 2: Konstruksi Skala

Pada Fase konstruksi skala dilaksanakan bertujuan untuk menyusun kesesuaian dimensi dan item dengan melakukan CFA (*Confirmatory Factor Analysis*) berdasarkan data empiris di lapangan.

2.1. Partisipan dan Prosedur Sampling

Populasi Target. Populasi target penelitian ini adalah mahasiswa S1 yang aktif dalam organisasi kemahasiswaan di Universitas X. Kriteria inklusi adalah: (a) mahasiswa S1 aktif (semester 2-8); (b) terdaftar sebagai anggota aktif atau pengurus organisasi kemahasiswaan (organisasi tingkat fakultas, universitas, atau Unit Kegiatan Mahasiswa); dan (c) bersedia berpartisipasi secara sukarela. Kriteria eksklusi adalah: (a) mahasiswa yang sedang cuti akademik; dan (b) mahasiswa yang tidak melengkapi seluruh item dalam instrumen.

Metode Sampling. Penelitian ini menggunakan *convenience sampling* dengan *stratified approach* berdasarkan fakultas untuk memastikan representasi dari berbagai disiplin ilmu. Data enumerator (mahasiswa yang dilatih untuk membantu distribusi instrumen) direkrut dari setiap fakultas untuk memfasilitasi akses ke responden.

Ukuran Sampel. Ukuran sampel ditentukan berdasarkan rekomendasi untuk *Confirmatory Factor Analysis* (CFA). Kline (2016) merekomendasikan minimal 10-20 responden per parameter yang diestimasi. Dengan 27 item dan 4 faktor laten, jumlah parameter yang diestimasi adalah sekitar 60 (27 *loading factors*, 27 *error variances*, 4 *factor variances*, 6 *factor covariances*). Dengan rasio konservatif 10:1, ukuran sampel minimal adalah 600. Namun, mengingat keterbatasan sumber daya dan waktu, target sampel ditetapkan 250-300 responden, yang masih memenuhi kriteria minimal 5:1 yang lebih longgar tetapi dapat diterima untuk CFA (Bentler & Chou, 1987).

Karakteristik Sampel. Sebanyak 268 mahasiswa berpartisipasi dalam penelitian ini. Setelah data *cleaning* (eliminasi responden dengan *missing data* > 10% atau pola respons yang tidak valid seperti *straight-lining*), 252 responden dipertahankan untuk analisis (*response rate* yang valid = 94.0%). Tabel 2 menyajikan karakteristik demografis sampel.

Fase 3: Evaluasi Skala

Fase Evaluasi bertujuan untuk menguji kualitas konstruk skala. Pada fase ini dilakukan uji model untuk membuktikan kesesuaian model skala.

3.1. Item Analysis

Item analysis dilakukan untuk mengevaluasi kualitas item sebelum CFA. Analisis meliputi:

Corrected Item-Total Correlation (CITC). CITC dihitung untuk setiap item dengan mengkorelasikan skor item dengan skor total skala (dengan item tersebut dikeluarkan). Kriteria yang digunakan adalah $CITC \geq 0.30$ (Field, 2013). Hasil menunjukkan bahwa dari 27 item, 14 item memiliki $CITC \geq 0.30$ (*range*: 0.331-0.555),

Tabel 2
Karakteristik Demografis Sampel (N = 252)

VARIABEL	KATEGORI	FREKUENSI N=252	PERSENTASE
USIA	18 - 20 tahun	126	50%
	21 - 24 tahun	123	49%
	26 tahun keatas	3	01%
JENIS KELAMIN	Laki - laki	110	44%
	Perempuan	142	56%
ASAL DOMISILI	Desa	122	48%
	Kota	130	52%
PRODI	Fakultas Ekonomi	23	9,13%
	Fakultas Syariah	76	30%
	Fakultas Sains	18	7,14%
	Fakultas Tarbiyah	20	8%
	Fakultas Psikologi	103	41%
	Fakultas Humaniora	12	4,76%

sedangkan 13 item memiliki $CITC < 0.30$. Item dengan CITC rendah dieliminasi karena mengindikasikan kontribusi yang lemah terhadap konstruk total.

Cronbach's Alpha if Item Deleted. Analisis ini menunjukkan apakah eliminasi item tertentu akan meningkatkan reliabilitas skala. Tidak ada item yang, jika dihapus, akan meningkatkan *alpha* secara substansial ($\Delta\alpha > 0.05$).

Berdasarkan *item analysis*, 14 item dipertahankan untuk analisis CFA. Distribusi item final adalah: *Pressure* (3 item: F01, F03, F06), *Rationalization* (3 item: F08, F13, F14), *Opportunity* (5 item: F15, F16, F17, F20, F21), dan *Capability* (3 item: F22, F24, F25).

3.2. Confirmatory Factor Analysis (CFA)

CFA dilakukan untuk menguji apakah struktur empat faktor yang dihipotesiskan (berdasarkan *Fraud Diamond Theory*) sesuai dengan data. Model yang diuji adalah *first-order CFA* dengan empat faktor laten (*Pressure*, *Opportunity*, *Rationalization*, *Capability*) yang berkorelasi satu sama lain. Metode Estimasi. *Maximum Likelihood with robust standard errors* (MLR) digunakan karena pelanggaran ringan terhadap normalitas multivariat. MLR memberikan estimasi parameter yang *robust* dan *standard errors* yang disesuaikan untuk non-normalitas (Kline, 2016).

Fit Indices. Kesesuaian model dievaluasi menggunakan multiple fit indices sebagaimana direkomendasikan oleh Hu dan Bentler (1999) dan Kline (2016): Chi-square (χ^2): Uji signifikansi statistik untuk model fit. Nilai non-signifikan ($p >$

.05) mengindikasikan fit yang baik, tetapi χ^2 sensitif terhadap ukuran sampel. *Relative Chi-square* (χ^2/df): Rasio χ^2 terhadap *degrees of freedom*. Kriteria: < 3 = fit yang baik; < 5 = fit yang dapat diterima. *Comparative Fit Index* (CFI): Membandingkan model yang diuji dengan *independence model*. Kriteria: ≥ 0.95 = fit yang baik; ≥ 0.90 = fit yang dapat diterima. *Tucker-Lewis Index* (TLI): Serupa dengan CFI tetapi memperhitungkan parsimoni model. Kriteria: ≥ 0.95 = fit yang baik; ≥ 0.90 = fit yang dapat diterima. *Root Mean Square Error of Approximation* (RMSEA): Mengukur *discrepancy per degree of freedom*. Kriteria: < 0.05 = fit yang baik; < 0.08 = fit yang dapat diterima. *Standardized Root Mean Square Residual* (SRMR): Rata-rata *standardized residual*. Kriteria: < 0.05 = fit yang baik; < 0.08 = fit yang dapat diterima. *Goodness of Fit Index* (GFI) dan *Adjusted Goodness of Fit Index* (AGFI): Proporsi varians yang dijelaskan oleh model. Kriteria: ≥ 0.95 = fit yang baik; ≥ 0.90 = fit yang dapat diterima.

3.3. Reliabilitas

Reliabilitas internal dievaluasi menggunakan dua metode:

Pertama, *Cronbach's Alpha* (α). Alpha dihitung untuk skala total dan untuk setiap subdimensi. Kriteria: $\alpha \geq 0.70$ = *acceptable*; $\alpha \geq 0.80$ = *good*; $\alpha \geq 0.90$ = *excellent* (Nunnally & Bernstein, 1994). Kedua, *Composite Reliability* (CR). CR dihitung untuk setiap faktor laten berdasarkan hasil CFA menggunakan formula: $CR = (\sum \lambda)^2 / [(\sum \lambda)^2 + \sum \delta]$ di mana λ adalah *standardized loading factors* dan δ adalah *error variances*. CR dianggap lebih akurat daripada Cronbach's alpha dalam konteks CFA karena memperhitungkan muatan faktor aktual (Hair et al., 2019). Kriteria: $CR \geq 0.70$ = *acceptable*; $CR \geq 0.80$ = *good*.

Tabel 3

Hasil Analisis Keseuaian Model

Uji Kecocokan	Kriteria	Hasil	Keterangan
CMIN/DF	CMIN /DF ≤ 3	1.67	Good Fit
GFI	AGFI $> 0,90$	0.952	Good Fit
AGFI	AGFI $> 0,90$	0.929	Good Fit
RMSEA	RMSEA $< 0,08$	0.038	Good Fit
RMR	RMR $< 0,08$	0.028	Good Fit

3.4. Validitas

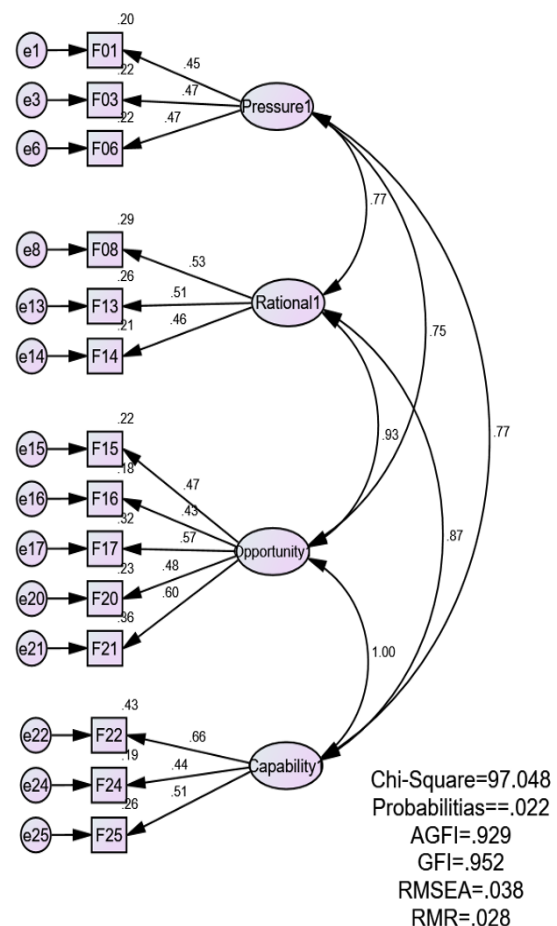
Construct Validity. *Construct validity* dievaluasi melalui CFA, dengan fokus pada apakah struktur empat faktor sesuai dengan data (confirmatory fit).

3.5. Software Analisis

Analisis data dilakukan menggunakan: IBM SPSS Statistics 26 untuk *data cleaning*, statistik deskriptif, *item analysis*, dan *Cronbach's alpha*. IBM AMOS 26 untuk *Confirmatory Factor Analysis*.

HASIL DAN PEMBAHASAN

Dari item yang diujicobakan dalam penelitian ini menunjukkan hasil ada 14 item yang mempunyai nilai diatas standar 0.3. dengan rentan skor 0.331-0.555. dengan *alpha cronbach* sebesar 0.8033. Kemudian dilakukan uji CFA dengan



yang menunjukkan hasil sebagai berikut:

Hasil konfirmatori analisis menunjukkan bahwa semua parameter kesesuaian model mengindikasikan hasil fit, artinya bahwa pada masing-masing aspek memiliki indikasi yang baik. Dengan demikian skala ini menunjukkan hasil yang baik.

Adapun *loading factor* pada masing-masing item menunjukkan hasil sebagai berikut:

Tabel 4
Loading Factor

			Estimate
F01	<---	Pressure1	.450
F03	<---	Pressure1	.472
F06	<---	Pressure1	.469
F08	<---	Rational1	.535
F13	<---	Rational1	.507
F14	<---	Rational1	.463
F15	<---	Opportunity1	.466
F16	<---	Opportunity1	.427
F17	<---	Opportunity1	.569
F20	<---	Opportunity1	.484
F21	<---	Opportunity1	.600
F22	<---	Capability1	.656
F24	<---	Capability1	.439
F25	<---	Capability1	.506

Loading factor menunjukkan hasil yang cukup baik, semua item mempunyai nilai diatas 0,4.

Penelitian ini bertujuan untuk mengembangkan dan memvalidasi *Fraud Risk Profiling Scale* (FRPS) berbasis *Fraud Diamond Theory* dengan format *Situational Judgment Test* untuk populasi mahasiswa. Hasil analisis menunjukkan bahwa FRPS memiliki properti psikometrik yang menjanjikan sebagai instrumen skrining awal, meskipun terdapat beberapa area yang memerlukan perbaikan.

Temuan utama penelitian ini adalah: (1) Struktur empat dimensi *Fraud Diamond* (*Pressure*, *Opportunity*, *Rationalization*, *Capability*) terkonfirmasi dengan sangat baik melalui CFA, dengan *fit indices* yang *excellent* ($\chi^2/df = 1.367$, CFI = 0.962, TLI = 0.952, RMSEA = 0.038); (2) Reliabilitas internal untuk skor total sangat baik ($\alpha = 0.803$) dan *acceptable* untuk subdimensi ($\alpha = 0.682-0.742$); (3) *Discriminant validity* antar dimensi sangat baik, dengan korelasi antar faktor yang rendah ($r = 0.113-0.195$); (4) Namun, muatan faktor (*factor loadings*) untuk sebagian besar item relatif rendah (57.1% item memiliki loading < 0.50), mengindikasikan bahwa beberapa item memerlukan revisi atau eliminasi; dan (5) *Convergent validity marginal*, dengan AVE untuk semua dimensi < 0.50 tetapi > 0.38.

Interpretasi Temuan dalam Konteks Teoretis

Konfirmasi Struktur *Fraud Diamond*. Temuan bahwa struktur empat faktor sesuai dengan data memberikan dukungan empiris yang kuat untuk *Fraud Diamond Theory* dalam konteks mahasiswa Indonesia. *Fit indices* yang *excellent* mengindikasikan bahwa keempat dimensi, *Pressure*, *Opportunity*, *Rationalization*, dan *Capability*, adalah komponen yang distinct tetapi saling terkait dalam menjelaskan risiko *fraud*. Ini konsisten dengan proposisi Wolfe dan Hermanson (2004) bahwa *fraud* memerlukan konvergensi dari keempat elemen ini.

Namun, korelasi antar dimensi yang sangat rendah (rata-rata $r = 0.145$) menimbulkan pertanyaan teoretis yang menarik. Di satu sisi, korelasi rendah mengindikasikan *discriminant*

validity yang baik keempat dimensi mengukur aspek yang berbeda dari *fraud risk*. Di sisi lain, korelasi yang sangat rendah dapat mengindikasikan bahwa keempat dimensi mungkin tidak sepenuhnya merupakan manifestasi dari satu konstruk umum (*fraud risk*), tetapi mungkin adalah konstruk yang relatif independen yang berkontribusi pada *fraud* melalui mekanisme yang berbeda. Interpretasi alternatif adalah bahwa *fraud risk* mungkin lebih tepat dikonseptualisasikan sebagai *formative construct* daripada *reflective construct* (Coltman et al., 2008). Dalam model *reflective*, indikator (item) adalah manifestasi dari konstruk laten, dan oleh karena itu harus berkorelasi tinggi. Dalam model *formative*, indikator membentuk atau menyebabkan konstruk, dan tidak harus berkorelasi tinggi satu sama lain. *Fraud risk* mungkin adalah *formative construct* di mana *Pressure*, *Opportunity*, *Rationalization*, dan *Capability* adalah komponen yang berbeda yang bersama-sama membentuk risiko *fraud*, tetapi tidak harus terjadi bersamaan pada setiap individu. Penelitian lanjutan dengan pendekatan *formative modeling* mungkin diperlukan untuk menguji hipotesis ini.

Peran Konteks Budaya

Penelitian ini dilakukan dalam konteks mahasiswa Indonesia, yang memiliki karakteristik budaya yang unik. Indonesia memiliki skor tinggi pada dimensi kolektivisme dan power distance dalam kerangka Hofstede (Hofstede Insights, 2024). Budaya kolektifis mungkin mempengaruhi manifestasi *fraud risk*, terutama dalam dimensi *Rationalization*. Dalam budaya kolektifis, individu mungkin lebih mudah merasionalisasi perilaku *fraud* jika tindakan tersebut dianggap menguntungkan kelompok (organisasi, teman) daripada diri sendiri. Ini tercermin dalam hasil wawancara kualitatif, di mana responden sering merasionalisasi perilaku *questionable* dengan alasan "untuk kepentingan organisasi" atau "semua organisasi juga begitu". *Power distance* yang tinggi mungkin juga mempengaruhi dimensi *Capability*. Dalam budaya dengan *power distance* tinggi, posisi atau jabatan memberikan otoritas yang lebih besar dan kurang dipertanyakan. Mahasiswa yang menjabat sebagai ketua atau bendahara organisasi mungkin memiliki akses dan otoritas yang lebih besar untuk melakukan *fraud* dengan sedikit pengawasan. Penelitian *cross-cultural* diperlukan untuk menguji apakah struktur dan manifestasi *fraud risk* berbeda antara budaya individualis dan kolektifis.

Perbandingan dengan Literatur

Perbandingan Fit Indices. Fit indices yang diperoleh dalam penelitian ini (CFI = 0.962, TLI = 0.952, RMSEA = 0.038) sangat baik dan sebanding atau bahkan lebih baik daripada studi pengembangan skala *fraud* lainnya. Misalnya, Julian et al. (2022) melaporkan CFI = 0.92 dan RMSEA = 0.06 untuk *Fraud Risk Judgment Scale*. Veríssimo et al. (2025) melaporkan CFI = 0.94 dan RMSEA = 0.05 untuk *Academic Misconduct Questionnaire*. Fit indices yang *excellent* dalam penelitian ini mengindikasikan bahwa struktur teoretis *Fraud Diamond* sangat sesuai dengan data empiris, memberikan validitas konstruk yang kuat. Perbandingan Reliabilitas. Reliabilitas total ($\alpha = 0.803$) dalam penelitian ini comparable dengan instrumen *fraud* lainnya. Julian et al. (2022) melaporkan $\alpha = 0.87$ untuk *Fraud Risk Judgment Scale* (15 item). Veríssimo et al. (2025) melaporkan $\alpha = 0.82$ untuk *Academic Misconduct Questionnaire* (18 item). Dengan hanya 14 item, FRPS mencapai reliabilitas yang sebanding, mengindikasikan efisiensi instrumen yang baik. Namun, reliabilitas subdimensi ($\alpha = 0.682-0.742$) sedikit lebih rendah, yang dapat dipahami mengingat jumlah item per dimensi yang sedikit (3-5 item). Penelitian menunjukkan bahwa reliabilitas cenderung lebih rendah untuk skala dengan jumlah item yang sedikit, dan $\alpha = 0.68-0.74$ dianggap *acceptable* dalam konteks ini (Cortina, 1993).

Perbandingan Factor Loadings

Dalam hal ini seputar area di mana FRPS menunjukkan kelemahan dibandingkan dengan instrumen lain. Rata-rata *factor loading* dalam penelitian ini adalah 0.506, dengan 57.1% item memiliki *loading* < 0.50. Sebagai perbandingan, Julian et al. (2022) melaporkan rata-rata *loading* 0.68 (range: 0.58-0.82), dan Veríssimo et al. (2025) melaporkan rata-rata *loading* 0.64 (range: 0.52-0.78). *Loading* yang lebih rendah dalam penelitian ini mungkin disebabkan oleh beberapa faktor yang akan didiskusikan di bawah.

Factor loadings yang rendah adalah kelemahan paling signifikan dalam penelitian ini dan memerlukan diskusi yang mendalam. Terdapat beberapa penjelasan potensial untuk *loading* yang rendah:

Pertama, Karakteristik Format SJT. SJT *inherently* memiliki kompleksitas yang lebih tinggi dibandingkan format *Likert scale* tradisional. Setiap item SJT adalah skenario yang kompleks dengan *multiple response options*, dan responden harus memproses informasi yang lebih banyak. Kompleksitas ini dapat menyebabkan *varians* yang lebih besar dalam respons, yang pada gilirannya dapat menurunkan

loading factors (Whetzel et al., 2020). Penelitian menunjukkan bahwa SJT umumnya memiliki *loading factors* yang lebih rendah dibandingkan dengan skala Likert untuk konstruk yang sama (Christian et al., 2010). Oleh karena itu, *loading* 0.40-0.50 dalam SJT mungkin setara dengan *loading* 0.60-0.70 dalam skala Likert dalam hal kontribusi terhadap konstruk.

Kedua, Redaksional Item yang kurang optimal. Beberapa item mungkin memiliki redaksional yang kurang jelas atau pilihan respons yang ambigu, menyebabkan responden menginterpretasikan skenario secara berbeda. Misalnya, item F16 (*Opportunity*) yang memiliki *loading* terendah (0.427) mungkin memiliki skenario yang terlalu kompleks atau pilihan respons yang tidak cukup *distinct*. Analisis kualitatif terhadap item dengan *loading* rendah diperlukan untuk mengidentifikasi masalah redaksional spesifik.

Ketiga, tumpang tindih atau ambiguitas dimensi. Meskipun korelasi antar dimensi rendah, beberapa item mungkin secara konseptual tumpang tindih antara dua dimensi. Misalnya, item yang dirancang untuk mengukur *Opportunity* mungkin juga melibatkan elemen *Rationalization* jika skenario menggambarkan situasi di mana kesempatan ada karena norma sosial yang *permissive*. Tumpang tindih ini dapat menyebabkan item memiliki *loading* yang terbagi (*split loading*) atau *loading* yang lebih rendah pada dimensi yang dituju.

Keempat, karakteristik sampel. Sampel penelitian ini terdiri dari mahasiswa dari satu universitas, yang mungkin memiliki karakteristik atau pengalaman yang unik. Jika skenario dalam beberapa item tidak relevan atau tidak familiar bagi sebagian besar responden, mereka mungkin merespons secara *random* atau berdasarkan *social desirability*, bukan berdasarkan kecenderungan *fraud risk* mereka yang sebenarnya. Ini dapat meningkatkan *varians error* dan menurunkan *loading*.

Kelima, *scoring key* yang subjektif. Meskipun *expert consensus* digunakan untuk menentukan *scoring key*, tetap ada kemungkinan bahwa *scoring* yang ditentukan oleh ahli tidak sepenuhnya mencerminkan bagaimana responden sebenarnya menilai kesesuaian respons. Penelitian menunjukkan bahwa *scoring key* yang berbeda (*expert consensus vs empirical keying*) dapat menghasilkan properti psikometrik yang berbeda (De Leng et al., 2017). Penelitian lanjutan dengan *empirical keying* (menggunakan kelompok kriteria seperti mahasiswa dengan vs tanpa riwayat pelanggaran) mungkin menghasilkan *loading* yang lebih tinggi.

Implikasi dari *loading* yang rendah. Meskipun *loading* yang rendah adalah suatu kelemahan, hal ini tidak sepenuhnya meniadakan utilitas FRPS. Pertama, *fit indices* yang *excellent* mengindikasikan bahwa struktur model secara keseluruhan sangat baik, meskipun kontribusi individual item lemah. Kedua, reliabilitas total yang baik ($\alpha = 0.803$) mengindikasikan bahwa instrumen secara keseluruhan konsisten dalam mengukur konstruk. Ketiga, untuk tujuan skrining awal (bukan diagnosis), instrumen dengan *loading* yang lebih rendah tetapi struktur yang jelas masih dapat berguna.

Namun, untuk meningkatkan kualitas instrumen, beberapa langkah perbaikan sangat direkomendasikan: (1) Revisi item dengan *loading* < 0.45 (F16, F24) berdasarkan analisis kualitatif dan *feedback* dari responden; (2) Pengembangan item baru untuk menggantikan item dengan *loading* rendah; (3) Pilot testing yang lebih ekstensif dengan *cognitive interviewing* untuk memastikan bahwa responden memahami skenario dan pilihan respons dengan cara yang dimaksudkan; dan (4) Validasi dengan sampel yang lebih besar dan beragam untuk menguji stabilitas *loading factors*.

Implikasi Teoretis

Penelitian ini memberikan beberapa kontribusi teoretis penting. Pertama, penelitian ini adalah yang pertama mengoperasionalkan keempat dimensi *Fraud Diamond* secara komprehensif dalam satu instrumen pengukuran untuk populasi mahasiswa. Sebagian besar penelitian sebelumnya menggunakan *Fraud Diamond* sebagai kerangka analisis *post-hoc* untuk menjelaskan kasus *fraud* yang sudah terjadi, bukan sebagai dasar untuk pengembangan instrumen prediktif. Penelitian ini menunjukkan bahwa *Fraud Diamond* dapat dioperasionalkan sebagai instrumen pengukuran dengan validitas konstruk yang baik. Kedua, penelitian ini berkontribusi pada literatur tentang SJT dengan menunjukkan aplikasi format ini dalam konteks pengukuran *fraud risk*. Sebagian besar penelitian SJT fokus pada pengukuran kompetensi atau kinerja dalam konteks seleksi karyawan atau penilaian profesionalisme dalam pendidikan kesehatan (Whetzel et al., 2020; Smith et al., 2020). Penelitian ini menunjukkan bahwa SJT juga dapat digunakan untuk mengukur konstruk yang lebih sensitif seperti *fraud risk*, dengan keunggulan mengurangi bias *social desirability*.

Ketiga, penelitian ini menimbulkan pertanyaan teoretis tentang struktur *fraud risk*: apakah *fraud risk* adalah reflective construct (di mana keempat dimensi adalah manifestasi dari konstruk laten

tunggal) atau *formative construct* (di mana keempat dimensi adalah komponen independen yang bersama-sama membentuk risiko)? Korelasi yang sangat rendah antar dimensi ($r = 0.113-0.195$) menunjukkan bahwa model *formative* mungkin lebih tepat. Penelitian lanjutan dengan pendekatan *formative modeling* (misalnya, menggunakan *Partial Least Squares Structural Equation Modeling*) diperlukan untuk menguji hipotesis ini.

Implikasi Praktis

Fraud Risk Profiling Scale (FRPS) memiliki potensi besar sebagai alat skrining di lingkungan pendidikan tinggi untuk mengidentifikasi mahasiswa yang berpotensi melakukan kecurangan, baik dalam konteks akademik maupun organisasional. Salah satu aplikasinya adalah dalam proses rekrutmen pengurus organisasi kemahasiswaan, terutama yang mengelola dana besar. FRPS dapat membantu menyeleksi calon pengurus, seperti bendahara atau ketua, dengan risiko *fraud* yang lebih rendah, atau sebaliknya, mengidentifikasi mereka yang perlu mendapat perhatian khusus, seperti pelatihan integritas atau pengawasan ekstra.

FRPS juga berguna untuk mengidentifikasi kebutuhan intervensi preventif yang personal. Misalnya, mahasiswa dengan tekanan finansial tinggi (skor tinggi pada dimensi *Pressure*) bisa ditawarkan bantuan konseling atau dukungan ekonomi, sementara yang cenderung membenarkan kecurangan (*Rationalization*) dapat diberi program pembangunan karakter dan pendidikan etika. Pendekatan seperti ini lebih efektif dibandingkan strategi seragam untuk semua. FRPS juga dapat digunakan sebagai instrumen evaluasi: institusi bisa mengukur efektivitas program anti-*fraud* atau pendidikan integritas dengan membandingkan skor FRPS sebelum dan sesudah intervensi. Penurunan risiko *fraud* setelah program berjalan bisa menjadi indikator keberhasilan. Lebih jauh, skala ini membuka peluang bagi penelitian lanjutan tentang faktor-faktor yang memengaruhi kecenderungan *fraud*, seperti latar belakang sosial-ekonomi, prestasi akademik, keterlibatan organisasi, atau ciri kepribadian tertentu. Namun, penggunaan FRPS harus tetap memperhatikan aspek etis. Skala ini hanya mengukur kecenderungan, bukan perilaku *actual*, artinya, skor tinggi tidak serta merta menandakan seseorang akan melakukan kecurangan. Oleh karena itu, FRPS tidak boleh menjadi satu-satunya dasar untuk menolak seseorang dari posisi tertentu. Transparansi, persetujuan informasi (*informed consent*), dan kerahasiaan

data individu harus dijaga agar hasil tes tidak disalahgunakan atau menyebabkan stigmatisasi.

Keterbatasan Penelitian

Penelitian ini memiliki sejumlah keterbatasan yang perlu diperhatikan dalam menafsirkan hasilnya. Pertama, sampel penelitian hanya berasal dari satu universitas (Universitas X) dan dikumpulkan melalui *convenience sampling*, sehingga tidak mewakili keragaman mahasiswa di berbagai perguruan tinggi, apalagi populasi non-mahasiswa. Karakteristik institusi seperti status negeri/swasta atau lokasi geografis bisa memengaruhi cara *fraud risk* muncul, sehingga generalisasi temuan harus dilakukan dengan hati-hati. Kedua, penelitian ini belum menguji validitas eksternal secara memadai; meskipun validitas konstruk telah dibuktikan melalui *Confirmatory Factor Analysis* (CFA), belum ada bukti *convergent*, *discriminant*, maupun *criterion validity*, misalnya, korelasi dengan perilaku *fraud* aktual atau instrumen serupa lainnya. Tanpa validasi eksternal, kepastian bahwa FRPS benar-benar mengukur risiko kecurangan masih terbatas. Ketiga, uji coba awal (*pilot test*) dilakukan pada kelompok kecil ($n = 30$) tanpa pendekatan kualitatif seperti *cognitive interviewing*, yang sebenarnya bisa membantu memahami bagaimana responden menafsirkan item-item skala. Keempat, desain *cross-sectional*, pengumpulan data hanya sekali waktu, menghambat evaluasi stabilitas skor seiring waktu (*test-retest reliability*) maupun kemampuan prediksi terhadap perilaku *fraud* di masa depan. Kelima, kunci penskoran (*scoring key*) dibuat berdasarkan konsensus ahli, yang meskipun cukup konsisten ($ICC = 0.82$), tetap bersifat subjektif. Pendekatan berbasis data empiris (*empirical keying*) bisa memberikan dasar yang lebih objektif. Keenam, karena seluruh data dikumpulkan melalui *self-report* dalam satu waktu, risiko bias *common method* termasuk kecenderungan menjawab sesuai norma sosial yang tidak dapat dihindari sepenuhnya. Ketujuh, penelitian ini langsung menggunakan CFA tanpa didahului *Exploratory Factor Analysis* (EFA), padahal praktik terbaik dalam pengembangan skala merekomendasikan EFA terlebih dahulu untuk mengeksplorasi struktur faktor secara empiris. Terakhir, banyak item (lebih dari separuhnya) memiliki *factor loading* di bawah 0.50, yang melemahkan validitas konvergen dan menunjukkan perlunya revisi atau penghapusan item tertentu.

Mengingat keterbatasan tersebut, beberapa langkah penting direkomendasikan untuk penelitian lanjutan. Pertama, FRPS perlu direvisi, terutama item dengan *loading* rendah (<0.45),

berdasarkan masukan kualitatif dari responden melalui *cognitive interviewing*, lalu divalidasi ulang pada sampel baru. Kedua, uji validitas eksternal harus dilakukan secara komprehensif: termasuk korelasi dengan skala integritas, etika, atau *Machiavellianism* (*convergent validity*); perbedaan dengan trait kepribadian umum seperti *Big Five* (*discriminant validity*); serta hubungan dengan riwayat pelanggaran atau penilaian pihak ketiga (*criterion validity*). Ketiga, studi longitudinal sangat diperlukan untuk menguji apakah skor FRPS benar-benar dapat memprediksi kecurangan di masa depan. Keempat, validasi harus diperluas ke sampel yang lebih beragam berbagai jenis perguruan tinggi, jenjang pendidikan, bahkan populasi non-mahasiswa seperti karyawan. Kelima, mengingat rendahnya korelasi antar dimensi *Fraud Diamond*, model *formative* (misalnya dengan PLS-SEM) layak dieksplorasi sebagai alternatif konseptualisasi *fraud risk*. Keenam, penelitian lintas budaya penting untuk memahami apakah struktur FRPS berlaku universal atau berbeda menurut konteks budaya. Ketujuh, untuk mendukung penerapan praktis, perlu dikembangkan norma dan *cut-off score* berdasarkan data representatif yang memungkinkan klasifikasi risiko (rendah/sedang/tinggi). Terakhir, FRPS dapat diintegrasikan ke dalam model prediktif yang lebih luas misalnya dengan variabel seperti *self-control*, tekanan finansial, atau status sosial dan dianalisis menggunakan pendekatan statistik canggih seperti *machine learning* untuk mengidentifikasi kombinasi faktor paling kuat dalam memprediksi kecurangan.

SIMPULAN

Penelitian ini berhasil mengembangkan dan memvalidasi *Fraud Risk Profiling Scale* (FRPS), sebuah instrumen pengukuran risiko *fraud* berbasis *Fraud Diamond Theory* dengan format *Situational Judgment Test* untuk populasi mahasiswa Indonesia. FRPS terdiri dari 14 item yang mengukur empat dimensi *fraud risk*: *Pressure*, *Opportunity*, *Rationalization*, dan *Capability*.

Hasil analisis menunjukkan bahwa FRPS memiliki properti psikometrik yang menjanjikan sebagai instrumen skrining awal. Struktur empat dimensi *Fraud Diamond* terkonfirmasi dengan sangat baik melalui *Confirmatory Factor Analysis*, dengan *fit indices* yang *excellent* ($\chi^2/df = 1.367$, CFI = 0.962, TLI = 0.952, RMSEA = 0.038, SRMR = 0.028). Reliabilitas internal untuk skor total sangat baik ($\alpha = 0.803$), dan *discriminant validity* antar dimensi sangat baik, dengan korelasi antar faktor yang rendah ($r = 0.113-0.195$).

Namun, penelitian ini juga mengidentifikasi beberapa kelemahan yang memerlukan perbaikan. Muatan faktor (*factor loadings*) untuk sebagian besar item relatif rendah (57.1% item memiliki *loading* < 0.50), mengindikasikan bahwa beberapa item memerlukan revisi atau eliminasi. *Convergent validity* juga marginal, dengan AVE untuk semua dimensi < 0.50. Selain itu, penelitian ini belum melakukan uji validitas eksternal (*convergent*, *discriminant*, dan *criterion validity*), yang merupakan keterbatasan mayor.

Meskipun demikian, FRPS merupakan langkah awal yang penting dalam pengembangan instrumen pengukuran *fraud risk* yang berbasis pada teori yang kuat dan menggunakan format yang dapat mengurangi *social desirability biases*. Dengan perbaikan yang direkomendasikan terutama revisi item dengan *loading* rendah dan uji validitas eksternal yang komprehensif FRPS memiliki potensi untuk menjadi alat yang berguna bagi institusi pendidikan tinggi dalam mengidentifikasi mahasiswa yang berisiko tinggi melakukan *fraud* dan mengembangkan intervensi preventif yang tepat sasaran.

Dari perspektif teoretis, penelitian ini memberikan dukungan empiris untuk *Fraud Diamond Theory* dalam konteks mahasiswa Indonesia dan menunjukkan bahwa keempat dimensi *fraud risk* dapat dioperasionisasikan sebagai instrumen pengukuran dengan validitas konstruk yang baik. Penelitian ini juga membuka jalan untuk penelitian lanjutan tentang struktur *fraud risk* (reflective vs formative), peran konteks budaya dalam manifestasi *fraud risk*, dan pengembangan model prediktif yang lebih komprehensif.

DAFTAR PUSTAKA

- Albrecht, W. S., Albrecht, C. O., Albrecht, C. C., & Zimbelman, M. F. (2019). *Fraud* examination (6th ed.). Cengage Learning.
- Amelia, R., & Setiawan, A. (2021). Academic dishonesty among Indonesian university students: Prevalence and predictors. *Journal of Academic Ethics*, 19(3), 287-302. <https://doi.org/10.1007/s10805-020-09382-x>
- Anand, V., Ashforth, B. E., & Joshi, M. (2015). Business as usual: The acceptance and perpetuation of corruption in organizations. *Academy of Management Executive*, 19(4), 9-23. <https://doi.org/10.5465/ame.2005.19417904>
- Association of Certified Fraud Examiners (ACFE). (2022). Report to the nations: 2022 global study on occupational *fraud* and abuse. ACFE.
- Aylott, L. M. E., Buszewicz, M., King, M., & Barley, E. A. (2023). Development and validation of a situational judgement test to assess professionalism in mental health services. *BMC*

- Medical Education*, 23, 158. <https://doi.org/10.1186/s12909-023-04143-2>
- Baz, W. E., Ruel, S., Baz, A. E., & Aldabbas, H. (2016). The role of capability in *fraud* risk assessment: A conceptual analysis. *Journal of Financial Crime*, 23(4), 842-856. <https://doi.org/10.1108/JFC-03-2015-0016>
- Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, 16(1), 78-117. <https://doi.org/10.1177/0049124187016001004>
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, 6, 149. <https://doi.org/10.3389/fpubh.2018.00149>
- Carpenter, S. (2018). Ten steps in scale development and reporting: A guide for researchers. *Communication Methods and Measures*, 12(1), 25-44. <https://doi.org/10.1080/19312458.2017.1396583>
- Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, 63(1), 83-117. <https://doi.org/10.1111/j.1744-6570.2009.01163.x>
- Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: *Two applications of formative measurement*. *Journal of Business Research*, 61(12), 1250-1262. <https://doi.org/10.1016/j.jbusres.2008.01.013>
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98-104. <https://doi.org/10.1037/0021-9010.78.1.98>
- Cressey, D. R. (1953). *Other people's money: A study in the social psychology of embezzlement*. Free Press.
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1(1), 16-29. <https://doi.org/10.1037/1082-989X.1.1.16>
- Darwis, H., & Samukri, R. R. (2024). The impact of *fraud* on financial statement *fraud*: Analysis of the moderating role of the audit committee. *International Journal of Multidisciplinary on Science, Management, and Economics*, 2(3), 104-118.
- De Leng, W. E., Stegers-Jager, K. M., Husbands, A., Dowell, J. S., Born, M. P., & Themmen, A. P. N. (2017). Scoring method of a Situational Judgment Test: Influence on internal consistency reliability, adverse impact and correlation with personality? *Advances in Health Sciences*

- Education, 22(2), 243-265. <https://doi.org/10.1007/s10459-016-9720-7>
- Dias-Oliveira, E., Bártoło, A., Monteiro, S., & Pereira, A. (2024). "It is no big deal!": Fraud Diamond theory as an explanatory model for students' fraudulent academic behavior. *SAGE Open*, 14(2), 1-15. <https://doi.org/10.1177/21582440241266091>
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). SAGE Publications.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50. <https://doi.org/10.1177/002224378101800104>
- Free, C. (2015). Looking through the fraud triangle: A review and call for new directions. *Meditari Accountancy Research*, 23(2), 175-196. <https://doi.org/10.1108/MEDAR-02-2015-0009>
- Fritz, T., González Cruz, H., Janke, S., & Narciss, S. (2024). How to best measure academic dishonesty in students: A systematic review of self-report assessment methods and psychometric quality. *European Journal of Psychological Assessment*, 40(4), 321-335. <https://doi.org/10.1027/1015-5759/a000861>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.
- Hofstede Insights. (2024). Country comparison: Indonesia. <https://www.hofstede-insights.com/country-comparison/indonesia/>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Huang, C. C., Wang, Y. M., Wu, T. W., & Wang, P. A. (2013). An empirical analysis of the antecedents and performance consequences of using the moodle platform. *International Journal of Information and Education Technology*, 3(2), 217-221. <https://doi.org/10.7763/IJiet.2013.V3.267>
- International Center for Academic Integrity. (2022). Academic integrity statistics. <https://academicintegrity.org/statistics/>
- Julian, L., Johari, R. J., Said, J., & Wondabio, L. S. (2022). Fraud risk judgment measurement scale development [Special issue]. *Journal of Governance & Regulation*, 11(1), 303-311. <https://doi.org/10.22495/jgrv11i1siart10>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.
- Komisi Pemberantasan Korupsi (KPK). (2023). Laporan tahunan KPK 2023. KPK.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Lievens, F., & Motowidlo, S. J. (2016). Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology*, 9(1), 3-22. <https://doi.org/10.1017/iop.2015.71>
- Lokanan, M. E. (2015). Challenges to the fraud triangle: Questions on its usefulness. *Accounting Forum*, 39(3), 201-224. <https://doi.org/10.1016/j.accfor.2015.05.002>
- McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, 60(1), 63-91. <https://doi.org/10.1111/j.1744-6570.2007.00065.x>
- Murphy, P. R., & Free, C. (2016). Broadening the fraud triangle: Instrumental climate and fraud. *Behavioral Research in Accounting*, 28(1), 41-56. <https://doi.org/10.2308/bria-51083>
- Nguyen, N. T., Biderman, M. D., & McDaniel, M. A. (2005). Effects of response instructions on faking a situational judgment test. *International Journal of Selection and Assessment*, 13(4), 250-260. <https://doi.org/10.1111/j.1468-2389.2005.00322.x>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Ones, D. S., & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance*, 11(2-3), 245-269. <https://doi.org/10.1080/08959285.1998.9668033>
- Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224-239). Guilford Press.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879-903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Rahman, A. F., Handayani, N., & Muslichah, I. (2025). Fraud diamond insights: Predictors of financial statement fraud in Indonesian companies. *Acta Commercii*, 25(1), 1-12. <https://doi.org/10.4102/ac.v25i1.1367>
- Reurink, A. (2018). Financial fraud: A literature review. *Journal of Economic Surveys*, 32(5), 1292-1325. <https://doi.org/10.1111/joes.12294>

Smith, K. J., Emerson, D. J., Boster, C. R., & Everly, G. S. (2020). Development and validation of a situational judgement test to assess professionalism. *American Journal of Pharmaceutical Education*, 84(7), ajpe7771. <https://doi.org/10.5688/ajpe7771>

Veríssimo, A. C., Domingues, I., & Dourado, M. (2025). Validation of the academic misconduct questionnaire. *Frontiers in Education*, 10, 1234567. <https://doi.org/10.3389/feduc.2025.1234567>

Whetzel, D. L., Sullivan, T. S., & McCloy, R. A. (2020). Situational judgment tests: An overview of development practices and psychometric characteristics. *Personnel Assessment and Decisions*, 6(1), 1-16. <https://doi.org/10.25035/pad.2020.01.001>

Wolfe, D. T., & Hermanson, D. R. (2004). The *fraud diamond*: Considering the four elements of *fraud*. *The CPA Journal*, 74(12), 38-42.