Jurnal Penelitian dan Evaluasi Pendidikan
Volume 29, No. 1, June 2025 (26-42)
Online: http://journal.uny.ac.id/index.php/jpep

Developing an instrument to measure students' anti-corruption character using modern test theory

Ema Rachmawati*; Siswanto Siswanto
Universitas Negeri Yogyakarta, Indonesia
*Corresponding Author. E-mail: ema_rachmawati@uny.ac.id

ARTICLE INFO

Article History
Submitted:
19 April 2025
Revised:
22 April 2025
Accepted:
22 April 2025

Keywords
instrument development; integrity; anti-corruption character; IRT

Scan Me:

ABSTRACT

This study aims to: 1) develop an instrument construct to measure anti-corruption character (integrity); 2) test the quality of the developed integrity instrument; and 3) describe the character profile of economics and accounting students at universities in Indonesia. The research adopts the instrument development model by Istiyono (2020), which consists of three main stages: planning, testing, and measurement. The planning stage covers the formulation of objectives, drafting of items, and preparation of scoring guidelines. Trials and measurements were conducted on students from economics, accounting, economics education, and accounting education programs across eight Indonesian universities. The trial involved 200 students, and the measurement phase involved 176 students. Content validity was analyzed using Aiken's V (ranging from 0.38 to 1.00), construct validity using Confirmatory Factor Analysis (CFA), and reliability through construct reliability estimates. The instrument, based on polytomous four-category responses, was analyzed using the Graded Response Model (GRM) from Item Response Theory (IRT), showing good model fit and precision across the trait range (-4.14 to 4.31). The final instrument consists of nine aspects: honesty, discipline, care, responsibility, hard work, simplicity, independence, courage, and justice, totaling 45 statement items. CFA confirmed unidimensionality aligned with the nine-aspect construct. The results show 87% of students are at high (23%) and very high (64%) levels of integrity, while only 13% fall in the moderate to very low categories. Among all aspects, honesty and responsibility scored lower than the others.

To cite this article (in APA style):
Rachmawati, E., & Siswanto, S. (2025). Developing an instrument to measure students' anti-corruption character using modern test theory. *Jurnal Penelitian dan Evaluasi Pendidikan, 29(1)*, 26-42 doi: https://doi.org/10.21831/pep.v29i1.84560

INTRODUCTION

Corruption remains a deeply rooted structural issue in many developing countries, including Indonesia. It undermines economic and political systems, weakens educational institutions, erodes democratic values, and hinders social equity (Okoye et al., 2024). Transparency International's Corruption Perception Index (CPI) indicates that Indonesia's corruption levels remain alarming (Pertiwi & Ainsworth, 2021; Dipierro & Rella, 2024). highlighting the need for a holistic approach beyond law enforcement to include preventive measures such as character education. Higher education institutions as part of the national education system have a strategic role in instilling the values of integrity and public ethics to the younger generation (Utamirohmahsari, 2024). Students as agents of change are expected to be pioneers in realising an anti-corruption culture in society (Dewantara et al., 2021). However, character education must go beyond lectures; it must be integrated into curricula, pedagogy, and campus culture (Junaidah et al., 2022; Nadir, 2024). Therefore, assessing students' anti-corruption character is essential to ensure the effectiveness of value-based education.

Anti-corruption character education in higher education, particularly for economics and accounting students, is pivotal to addressing systemic corruption in Indonesia, as evidenced by

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i1.84560

its 2021 Corruption Perception Index (CPI) rank of 96th (score 38/100), reflecting entrenched vulnerabilities in strategic sectors (Sanjaya & Trifena, 2023; Wulandari et al., 2024). This score not only shows the high level of corruption, but also hints at the vulnerability of strategic sectors such as finance, business, and government to corrupt practices. Despite its importance, the measurement of students' anti-corruption character to date still faces conceptual and methodological challenges (Ginanjar & Purnama, 2023). Yet, current assessment tools often rely on Classical Test Theory (CTT), which is limited by its dependence on sample characteristics and inability to capture individual variation across ability levels (Alordiah & Oji, 2024). In the context of character measurement, where values and attitudes are latent (not directly observable), more sophisticated and accurate measurement approaches are needed.

Item Response Theory (IRT), or Modern Test Theory, addresses these limitations. It provides item-level analysis, adaptive testing, and sample-independent measurements ideal for assessing complex, latent traits like anti-corruption character (Gyamfi & Acquaye, 2023). In addition, IRT allows ensures calibrated item difficulty and discrimination, enhancing measurement validity and reliability (Fierro Bósquez et al., 2025; Fitraynsyah & Hilmiyati, 2024). The development of IRT-based psychometric instruments has shown significant progress, especially in the context of educational assessment and psychological measurement (Chang et al., 2021; Lang & Tay, 2021). However, there are still very limited studies that systematically apply this approach in measuring students' anti-corruption character. This shows a gap in the scientific literature and educational practice, where anti-corruption character measurement instruments are still not developed with adequate methodological standards (Paranata, 2025). In fact, valid and reliable measurements are very important in the process of evaluating and improving character education programmes in higher education.

Anti-corruption character itself includes a series of values and attitudes such as honesty, responsibility, caring, discipline, courage, and justice, which must be measured operationally through clear and standardised indicators (Dewantara et al., 2021; Maulidi et al., 2024). A good measurement instrument must be able to accurately represent these dimensions and allow interpretation of results that can be used as a basis for making educational policy decisions, designing character strengthening programmes (Manggaberani & Putro, 2024), and developing higher education-based corruption prevention strategies (Lozano-Peña et al., 2021). Therefore, the development of IRT-based instruments in measuring students' anti-corruption character is a significant innovation in the context of character education in higher education. Through this approach, the instrument not only functions as an evaluation tool, but also as a diagnostic instrument capable of providing important information about individual character strengths and weaknesses (Dickerson et al., 2025). In addition, IRT-based measurement results can also support the development of differentiated and personalised learning, where educational interventions can be tailored to specific student character profiles.

In line with the urgency of strengthening anti-corruption character education in higher education, this research specifically aims to develop a valid and reliable instrument in measuring the integrity character of students (Bantam & Nur Zhafarina, 2022; Yusoff et al., 2023).This study aims to develop a valid, reliable IRT-based instrument to measure students' anti-corruption character. The first goal is to construct indicators that reflect integrity, including honesty, responsibility, consistency, and moral courage . The second goal is to evaluate item quality through IRT analysis, focusing on difficulty, discrimination, and model fit (Lim, 2024). The third is to map integrity profiles among economics and accounting students nationwide, providing insights into the academic environments influencing ethical behavior (Cerratto Pargman & McGrath, 2021). By developing this IRT-based instrument, the study contributes to measurable, objective, and sustainable character education. It lays the groundwork for further research, including cross-disciplinary comparisons, mixed-method investigations of psychosocial factors, and cross-country validation to enhance generalisability. Ultimately, this

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i1.84560

research advances educational psychology and public policy while offering a strategic tool for strengthening integrity in higher education.

RESEARCH METHOD

This study used a research and development approach by adapting the affective instrument development model refined through Istiyono's model (Istiyono, 2020; Istiyono et al., 2014). This model consists of three main stages that are carried out systematically, namely planning, trials, and measurement. Each stage includes a series of steps designed to ensure the validity and reliability of the instrument developed. This research aims to produce a measurement instrument for students' anti-corruption character that is in accordance with the principles of measurement in Modern Test Theory (Item Response Theory/IRT), with a focus on the population of students of Economics, Accounting, Economics Education, and Accounting Education study programmes in universities in Indonesia.

Research Design

This research adopts Edi Istiyono's model-based instrument development approach, which includes three main stages: planning, piloting, and measurement (Istiyono et al., 2014). Each stage is carried out systematically to produce a valid and reliable instrument for measuring anti-corruption values in students of Economics, Accounting, Economics Education, and Accounting Education study programmes in various universities in Indonesia. This model emphasises the integration of theory and empirical evidence in the development of measuring instruments based on the Item Response Theory (IRT) approach.

Planning Stage

This stage begins with the preparation of measurement objectives and the development of an instrument grid based on nine aspects of anti-corruption values: honesty, discipline, care, responsibility, hard work, simplicity, independence, courage, and justice. Each aspect is translated into a number of sub-aspects and indicators, which are then outlined in the form of 45 statement items. The instrument was designed in the form of a questionnaire with a four-choice Likert scale model, using tiered scoring from 1 to 4. The next step was expert judgement by two expert lecturers to ensure the suitability of the content of the items with the indicators, as well as the clarity of the language and structure of the items.

Trial Stage

The pilot test was conducted in two stages. The limited trial (readability) involved five students, while the main trial was conducted on 200 students from various universities in Indonesia. The purpose of the pilot test was to prove the validity and reliability of the instrument. Content validity was analysed using Aiken's V index based on the assessment of two experts. Meanwhile, construct validity was tested through Exploratory Factor Analysis (EFA) with statistical requirements such as KMO values, Bartlett's Test, communalities, and loading factors, following the recommendations of (Hair et al., 2019). Reliability estimates were calculated using Cronbach's alpha as well as composite reliability estimates model fit test.

Measurement Stage

The revised instrument after the pilot test was used to measure anti-corruption character in 176 students from 11 universities in Indonesia. Data collection was conducted online through Google Form. The data obtained were analysed using the IRT approach for four-category

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i1.84560

polytomous to assess item characteristics and measurement reliability at the individual level. The measurement results were interpreted into five levels based on the ideal mean and ideal standard deviation: very high, high, medium, low, and very low.

Table 1. Matrix of the Character Assessment Instrument

No.	Aspects	Indicators	Item Number
1	Honesty	1) State or express facts and feelings as they are	1, 2, 3
		2) Willingness to recognise one's own mistakes and the strengths of others	4, 5
		3) Doing tasks according to ability	6
2	Careness	1) Sensitive to the difficulties of others.	7, 8, 9
		2) Sensitive to damage to the physical environment	10, 11
		3) Sensitive to various deviant behaviours	12, 13
3	Independen	1) Design your own learning according to your goals	14
		2) Choose a strategy and then implement the plan	15
		3) Can choose their own learning resources	16
		4) Doing tasks independently	17
		5) Monitor his/her learning progress, evaluate the results and compare with certain standards	18
4	Discipline	1) On time	19, 20
		2) Comply with the rules or regulations of the joint / campus	21
5	Responsibility	1) Carry out individual tasks well	22
		2) Accept risks and actions taken	23
		3) Not blaming/accusing others without accurate evidence	24
		4) Returning borrowed items	25
		5) Take responsibility for the actions taken	26
		6) Keeping promises	27
		7) Not blaming others for one's own wrong actions	28
6	Determination	1) Work earnestly until the goal is achieved	29
		2) Not giving up easily in the face of problems	30
7	Modesty	1) Attitude and behaviour are not excessive	31, 32
		2) Spending wealth according to need	33
		3) Humble	34, 35
8	Courageous	1) Not afraid to face danger/ difficulties/ challenges	36, 37, 38
		2) Self-confidence	39, 40
9	Fairness	1) Does not discriminate	41, 42
		2) Neutral	43
		3) Not labelling someone in a negative context on something that is not yet clearly true	44, 45

Data Analysis Technique

Data analysis was conducted in stages to ensure the validity, reliability, and model fit of the instrument, based on Modern Test Theory. Content validity was assessed using Aiken's V index, based on evaluations from two expert lecturers in educational measurement. Items were rated on a four-point scale: very suitable, quite suitable, needs revision, and not suitable. The Aiken's V formula (Aiken, 1985), was applied to calculate content validity. Construct validity was evaluated through Confirmatory Factor Analysis (CFA) to verify alignment with the theoretical construct of nine anti-corruption values, using goodness-of-fit indices (Goretzko et al., 2024). Instrument reliability was assessed using construct reliability, with a threshold of ≥ 0.70 indicating acceptable reliability (Retnawati, 2016). Item Response Theory (IRT) analysis was applied to polytomous data with four response categories to estimate item discrimination and difficulty, and to evaluate how well items differentiate students' integrity levels.

FINDINGS AND DISCUSSION

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i1.84560

Developing Constructs

The construct of the character assessment instrument includes nine main aspects: honesty, caring, independence, discipline, responsibility, hard work, modesty, courage, and fairness. The instrument matrix is shown in Table 1, illustrating the distribution of items on the indicators of the nine aspects totalling 45 items. The use of 45 items in this matrix also considers the principle of parsimony in Modern Test Theory, where each item must function accurately and fairly for all test takers (Marianti et al., 2021).

Content validity

The scores obtained for the assessment instrument in the form of Likert scale assessment were analysed using Aiken V index analysis. The analysis results can be categorised as valid if they meet the Aiken V index coefficient limit. According to (Retnawati, 2014), content validity can be categorised into three based on the validity index: less valid (index ≤ 0.4), moderately valid (index 0.4-0.8), and highly valid (index > 0.8). Instruments that have high content validity indicate that each item in the instrument is in accordance with the concept being measured, so that the measurement results can be trusted. The distribution of the content validity of the instrument items is stated in Table 2.

Table 2. Instrument Content Validity Categories Based on Aiken's V Index

Category V Aiken	Value Range	Item Number
High	> 0.80	1, 3, 4, 5, 6, 8, 21
Medium	> 0.40 – 0.80	2, 7, 9 - 11, 14 - 20, 22-30, 32-45
Low	≤ 0.40	13, 31

Instrument quality

Confirmatory factor analysis CFA

The instrument is tested to determine how well its items reflect the intended latent construct. The loading factor quantifies the relationship between each observed variable (indicator) and the underlying latent factor. Ideally, high loading factor values are preferred to ensure that indicators meaningfully represent the measured construct. However, in some contexts, values as low as 0.40 are considered acceptable, particularly when supported by a sufficiently large sample size. According to Hair et al. (2019), a loading factor of 0.40 is statistically significant for a sample of 200 respondents. Therefore, the choice of threshold should align with the specific research context, sample size, and disciplinary standards to ensure robust and meaningful measurement outcomes.

The Figure 1 reflect the modified integrity instrument model. Modifications were made based on modification indices due to poor model fit in the initial stage, as indicated by a Chi-Square p-value of 0.000 (<0.05), signaling statistical inadequacy (Alqahtani, 2022). After theoretical and empirical adjustments, all items showed loading factors above 0.40 and t-values above 1.96, confirming strong construct representation and statistical significance. These results, consistent across all aspects from caring to fairness, indicate improved model fit and reinforce the construct validity of the revised instrument.

After the model modification process based on the results of modification indices at the previous stage, the results of the CFA analysis in Table 3 show that the integrity instrument model developed has fulfilled all goodness of fit criteria. This is reflected in the five model feasibility indices which are all in the "fit" category based on applicable statistical criteria. The Chi-Square p-value of 0.06 (> 0.05) indicates that there is no significant difference between the estimated model and the empirical data, so the model is declared suitable (Pada et al., 2018).

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i1.84560

Furthermore, the RMSEA value of 0.02 (<0.08) reflects that the model's approximation error to the population is very low. The CFI and TLI indices of 0.98 each, well above the minimum limit of 0.90, show that the model has a very good fit when compared to the null model. Finally, the SRMR value of 0.05 (<0.08) indicates that the difference between the observed covariance matrix and that predicted by the model is within acceptable tolerance limits.

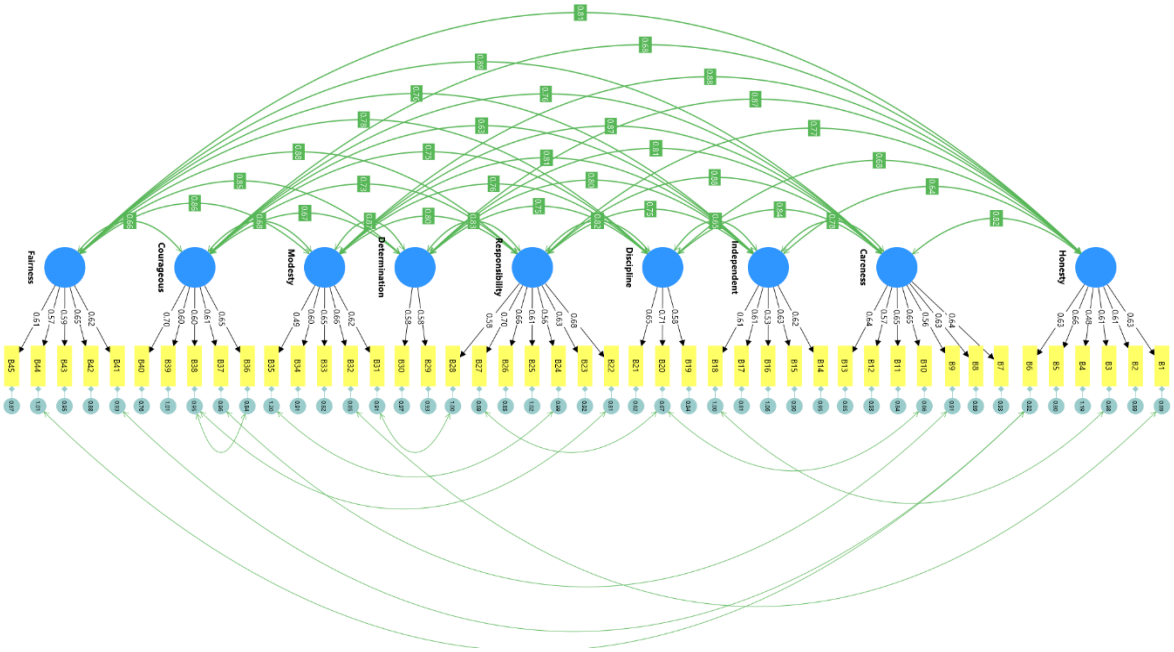

Figure 1. Path Diagram of 9 Aspects of Integrity

Table 3. Model Fit Criteria After Modification

No.	Fit Index	Output	Criteria	Description
1	*p-values* (Chi-Square)	0,06	≥ 0,05	Fit
2	Root Mean Square Error of Approximation (RMSEA)	0,02	< 0,08	Fit
3	Comparative Fit Index (CFI)	0,98	≥ 0,90	Fit
4	Tucker-Lewis Index (TLI)	0,98	≥ 0,90	Fit
5	Standardised Root Means Square (SRMR)	0,05	< 0,08	Fit

Composite Reliability

Construct reliability analysis for each set of questions is conducted to assess the extent to which each item contributes to the measured construct. The results of this analysis provide an idea of how effective the construct is in measuring the variable in question, as well as measuring the internal consistency of the items used in the model. According to Retnawati (2016), composite reliability is calculated using the formula:

$$CR = \frac{(\sum loading\ factor)^2}{(\sum loading\ factor)^2 + \sum(1 - i^2)} \quad (1)$$

From the table presented, the calculation of the total loading factor is obtained as 27.7, while the total error variance (∑(1 - i²)) is 27.85. Based on these calculations, the composite reliability value for the entire construct can be calculated as follows:

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i1.84560

$$CR = \frac{(27.7)^2}{(27.7)^2 + 27.85} = 0.96 \quad (2)$$

The value of CR = 0.96 greater than 0.80 indicates that this instrument has very high reliability, indicating that the internal consistency between indicators within each construct is very good (Catalán & Gordon, 2020). Therefore, this instrument can be said to have very high internal consistency in measuring the construct under this study.

IRT Assumptions

Unidimensional

The assumption of unidimensionality is a critical foundation to ensure that items in an instrument (such as an ability test) predominantly measure a single latent construct, in accordance with the basic principles of IRT. To evaluate this, Principal Component Analysis (PCA) is often used as an exploratory method to identify the underlying dimensional structure of the data (Gewers et al., 2022). The Table 4 presents the results of PCA analysis that illustrates the distribution of eigenvalues and variance contributions of each component.

Table 4. Total Variance Explained

Component	Initial eigenvalues		
	Total	% of Variance	Cumulative
Dim,1	14.543	32.318	32.3
Dim,2	1.714	3.809	36.1
Dim,3	1.595	3.545	39.7
Dim,4	1.549	3.442	43.1
Dim,5	1.255	2.788	45.9
Dim,6	1.216	2.701	48.6
Dim,7	1.188	2.640	51.2
Dim,8	1.095	2.433	53.7
Dim,9	1.070	2.378	56.1

Based on the Principal Component Analysis (PCA) results in the eigenvalue and total variance tables, the dimensionality structure of the data of 45 items shows characteristics relevant to the assumption of unidimensionality in Item Response Theory (IRT), showing an exponential decrease in the percentage of variance explained after the first component (Dim.1), with a percentage of 32.3% (eigenvalue = 14.543), far exceeding the next component (Dim.2 = 3.81%; eigenvalue = 1.714). This is also evidenced by the results of the scree plot used in the Principal Component Analysis (PCA), which illustrates the eigenvalues of a number of principal components extracted from the data as shown below.

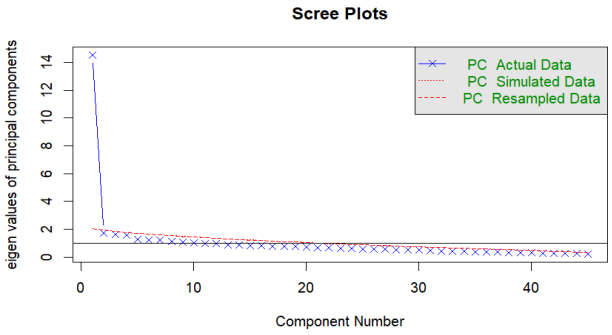

Figure 2. Scree Plots

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i1.84560

The Figure 2 provide a visualisation of how the eigenvalues decrease dramatically after the first component (elbow effect), indicating that the additional components are not substantively significant. Overall, the dominance of the first component remains strong evidence for unidimensionality in IRT.

Invariance test

Parameter invariance is essential to ensure consistent estimates of item discrimination (a), difficulty (b), and participant ability (theta) across groups or conditions (Luong & Flake, 2023). Discrimination (a) indicates how well an item differentiates based on latent ability, while difficulty (b) reflects the item's challenge level, and theta represents the individual's estimated ability (Germano et al., 2021). The table below shows the results of invariance testing, highlighting the equivalence of a, b, and theta values across groups, thus supporting the validity and measurement equivalence of the instrument.

Test for invariance of the difference power parameter (a) and parameter (b)

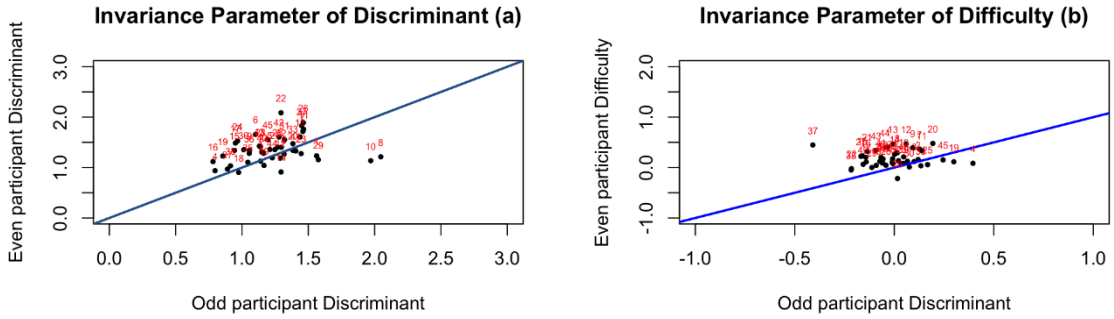

Figure 3. Test of Invariance of Variance Parameters (a**)**

Figure 3 visualises the results of the invariance test for the discrimination parameter (a) and difficulty parameter (b) across odd and even item subsets. The black dots represent item parameter estimates from both subsets, while the blue diagonal line indicates the identity line (y = x), reflecting perfect invariance. For parameter a, some items, such as item 10 and item 8, deviate noticeably from the line, suggesting potential inconsistencies in discrimination estimates. In contrast, parameter b estimates are mostly clustered around the origin, indicating low item difficulty and minimal variation between subsets. This suggests that item difficulty tends to be invariant and unaffected by how the data are partitioned.

Invariance test of ability parameter (θ)

The Figure 4 illustrates the invariance test of participants' ability parameters by comparing estimates from odd and even items. Each black dot represents individual ability estimates from both item sets, while the orange diagonal line (y = x) indicates perfect agreement. The close alignment of data points with this line and a high correlation of 0.916 reflect strong consistency between the two estimates. These findings demonstrate that the instrument meets the invariance assumption, confirming its stability and reliability in measuring ability across different item subsets within the IRT framework.

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i1.84560

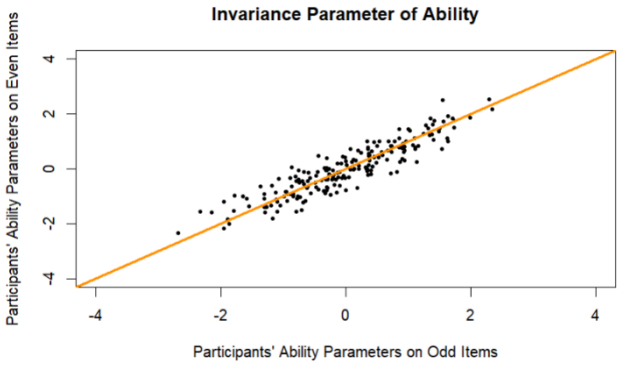

Figure 4. Ability Parameter Invariance Test (θ)

Model fit test

In Item Response Theory (IRT), model fit testing ensures that the selected model (e.g., Graded Response Model or Partial Credit Model) aligns with the data's response patterns. The M2 statistic is commonly used to compare observed and expected responses. To support this, additional methods such as LRT, AIC/BIC, and residual analysis are applied. The Table 5 summarizes the fit statistics used to evaluate the model's suitability.

Table 5. Results of Model Fit with M2 Statistic

	M2	df	p	RMSEA	RMSEA_5	RMSEA_95	SRMSR	TLI	CFI
GRM	1044	855	0.000009248	0.0333	0.0255	0.0401	0.0571	0.983	0.984

The model fit test in using the M2 statistic indicates that the Graded Response Model (GRM) yields a significant p-value ($p = 0.0000092$), suggesting a statistical difference between the empirical data and the model's expected responses. However, a significant p-value alone does not necessarily imply poor model fit, especially in large samples. Therefore, additional indices such as RMSEA (0.0333; 90% CI: 0.0255–0.0401), SRMSR (0.0571), TLI (0.983), and CFI (0.984) are used to provide a more comprehensive assessment. These values fall within acceptable ranges, indicating that the GRM fits the data well. Overall, these results support the model's validity and suggest that the GRM is a suitable choice for IRT-based measurement.

Test the suitability of each item

Item-level fit tests evaluate whether response patterns on polytomous items align with model expectations. Chi-square tests are commonly used to detect deviations, helping identify items that may need category revision, clarity improvements, or adjustments due to ability distribution. This ensures instrument integrity and enhances measurement quality.

The results of the item-by-item fit test provide an overview of how well each item in a polytomous instrument reflects the predictions of the IRT model used. Using Chi-square statistic to compare actual responses with model expectations, with interpretation based on p-value (significance value). A p-value > 0.05 indicates that there is no significant difference between the empirical data and the model predictions, so the item can be considered a good fit to the model. Of the 45 items analysed, the majority showed a good fit to the model, characterised by high p values that were well above the significance threshold. For example, items in the Honesty (B1-B6), Independence (B14-B18) and Responsibility (B22-B28) domains consistently showed p values > 0.25, reinforcing the internal validity of the instrument.

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i1.84560

Table 6. Fit Test Results for Each Item

Aspects	Number of Items	Suitable	Not Suitable
Honesty	6	6	0
Careness	7	6	1
Independent	5	5	0
Discipline	3	3	0
Responsibility	7	7	0
Determination	2	2	0
Modesty	5	3	2
Courageous	5	5	0
Fairness	5	5	0
Total	45	42	3

Item parameter estimation

The item parameter estimation results from the polytomous IRT model reveal how well each item discriminates test takers by ability (discrimination parameter, a) and the level of challenge posed by each item (difficulty parameter, b). Table 7 shows the value of the parameter a varies from 0.84 to 1.67, which generally reflects good to excellent discriminatory ability of the items. Items with higher a values such as B22 (a = 1.67) and B11 (a = 1.62) have higher sensitivity in discriminating participants at different ability levels, while items such as B16 (a = 0.94) or B37 (a = 0.93) show more moderate discriminatory ability. Meanwhile, the parameters b1, b2, and b3 on each item represent the transition points between response categories (thresholds), which conceptually mark the change in the probability of participants selecting a higher category in the response scale (e.g., from "agree" to "strongly agree"). Item centre locations (which are the means of the thresholds) generally ranged from 0.23 to 0.67, with the majority of items lying around the values of 0.4 to 0.5, indicating that item difficulty levels were in the region of average ability of participants.

Test information function

The Test Information Function (TIF) represents the combined information from all polytomous items in the instrument, indicating the test's precision in estimating participant ability (theta) across different levels. High TIF values show where the test is most accurate, while low values suggest areas needing item revision or addition. TIF analysis supports instrument optimisation by guiding improvements in measurement precision and coverage.

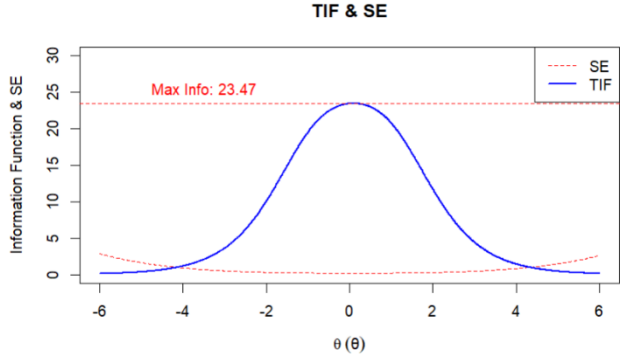

Figure 5. Test Information Function and Standard Error (SE)

The Figure 5shows the TIF retrived the maximum value of information is reached at the midpoint (around θ = 0) with a maximum TIF value of 23.47. This indicates that the test is most precise in measuring the ability of participants around the average ability. The purple dot markers at coordinates (-4.14, 1) and (4.31, 1) indicate that at both ends of the ability spectrum

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i1.84560

(very low and very high ability). Overall, this TIF analysis shows that the instrument is very efficient in measuring participants' overall ability with a low level of measurement error in the main ability range, and is quite robust at the extremes of ability.

Table 7. Item Parameter Estimates of Distinctiveness (a) and Difficulty (b)

Aspects	Item	a	b1	b2	b3	Location
Honesty	B1	1.38	-0.59	0.11	0.98	0.53
	B2	1.29	-0.72	0.09	1.06	0.50
	B3	1.11	-0.68	0.01	0.99	0.43
	B4	0.84	-0.75	0.30	1.17	0.48
	B5	1.28	-0.88	0.04	0.95	0.42
	B6	1.35	-0.65	0.15	0.80	0.46
Careness	B7	1.57	-0.38	0.26	0.86	0.61
	B8	1.55	-0.64	0.19	0.86	0.54
	B9	1.19	-0.81	0.30	1.25	0.58
	B10	1.44	-0.68	0.19	0.85	0.50
	B11	1.62	-0.49	0.26	0.96	0.63
	B12	1.23	-0.69	0.27	1.19	0.59
	B13	1.52	-0.55	0.26	1.01	0.61
Independent	B14	1.08	-0.74	0.16	0.93	0.43
	B15	1.14	-0.89	0.15	0.92	0.40
	B16	0.94	-0.96	0.05	1.08	0.38
	B17	1.22	-1.07	-0.04	1.15	0.42
	B18	0.94	-0.94	0.01	0.85	0.30
Discipline	B19	1.02	-0.88	0.19	1.26	0.50
	B20	1.39	-0.54	0.27	1.27	0.67
	B21	1.32	-0.78	0.13	0.94	0.47
Responsibility	B22	1.67	-0.78	-0.10	0.56	0.37
	B23	1.35	-0.71	0.08	0.80	0.43
	B24	1.23	-0.83	0.09	0.91	0.42
	B25	1.31	-0.59	0.19	0.74	0.45
	B26	1.60	-0.81	-0.11	0.62	0.36
	B27	1.65	-0.86	0.08	0.90	0.50
	B28	1.29	-0.76	-0.02	0.81	0.39
Determination	B29	1.32	-0.98	0.03	0.78	0.35
	B30	1.14	-0.83	-0.09	1.03	0.40
Modesty	B31	1.44	-0.68	0.13	0.82	0.47
	B32	1.35	-0.78	-0.02	0.86	0.41
	B33	1.41	-0.74	0.04	0.96	0.48
	B34	1.22	-0.81	0.03	1.03	0.45
	B35	1.07	-0.94	-0.11	0.65	0.23
Courageous	B36	1.18	-0.91	0.07	0.95	0.40
	B37	0.93	-1.01	0.17	0.98	0.36
	B38	0.97	-0.99	0.15	0.97	0.36
	B39	1.09	-0.76	0.09	0.85	0.38

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i1.84560

Fairness	B40	1.22	-0.81	0.14	0.85	0.41
	B41	1.37	-0.64	0.01	0.83	0.45
	B42	1.45	-0.77	-0.06	0.72	0.39
	B43	1.28	-0.74	0.17	0.99	0.49
	B44	1.21	-0.59	0.05	1.06	0.51
	B45	1.36	-0.7	0.18	1.11	0.56

Character profile

Table 8 shows the distribution of respondents across higher education institutions in Indonesia, providing an overview of the geographical and institutional spread. Analyzing this distribution is important to understand the sample representation in the study and its potential impact on the findings.

Table 8. Measurement Respondent Profile

No.	College/University	Frequency	Percentage (%)
1	University of Jember	2	1.1
2	University of Muhammadiyah Palembang	2	1.1
3	Pattimura University	1	0.6
4	University of Riau	1	0.6
5	Sanata Dharma University	1	0.6
6	UNPAD	6	3.4
7	UNY	104	59.1
8	UPI	13	7.4
9	UPS Tegal	29	16.5
10	USK	12	6.8
11	Unknown	5	2.8
Total		176	100

Estimation of Participant Integrity

Participant ability estimation in IRT allows educators and researchers to obtain a more accurate, objective and in-depth picture of an individual's ability level based on response patterns to a series of items. Unlike classical approaches that only rely on total scores, ability estimation in IRT takes into account item characteristics and participant abilities simultaneously, thus providing advantages in interpreting test results, designing adaptive instruments, and supporting more targeted decision-making in the context of educational assessment.

Based on Figure 6 of the distribution of participants' ability estimates, it can be seen that the majority of participants are in the middle ability range, i.e. in the intervals $-1 \leq \theta < 0$ and $0 \leq \theta < 1$. For all three estimation methods used Maximum Likelihood Estimation (MLE), Maximum a Posteriori (MAP), and Expected a Posteriori (EAP) these distributions show a consistent pattern, which reflects the stability and reliability of the approach in mapping participants' abilities. Meanwhile, the number of participants in the very low ($\theta < -4$) and very high ($\theta \geq 4$) ability categories was zero, indicating a rather normal distribution of the data without any extreme outliers. In the low ability range ($-4 \leq \theta < -2$), the number of participants was very small, ranging from 3 to 6 participants, confirming that only a small proportion of the population faced significant challenges in understanding the tested material.

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i1.84560

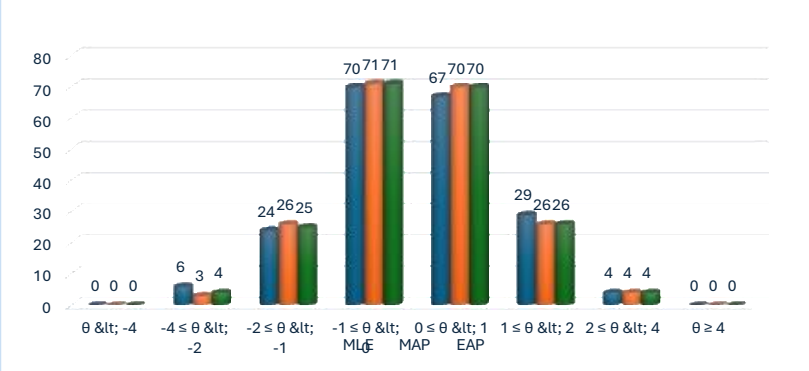

Figure 6. Estimated Ability of Participants

Similarly, participants with high ability ($2 \leq \theta < 4$) only numbered around 4 participants for each estimation method, indicating the existence of a group with superior academic potential but in a relatively small proportion. The estimation process applied has minimised bias and maximised the reliability of the results, so that it can be used as a strong basis for pedagogical and educational policy decision-making. Furthermore, the majority of participants concentrated at the middle ability level, it is advisable to develop items that are more varied in difficulty levels so that the instrument is able to differentiate more sharply between participants from across the ability spectrum. This approach not only improves the accuracy of individual ability estimation, but also enriches the overall quality of the assessment, in line with IRT principles in the development measurement tools.

CONCLUSION

The research conclusions show that: (1) This study designed an anti-corruption character measurement instrument based on Modern Test Theory (IRT), which includes nine aspects of integrity: honesty, discipline, caring, responsibility, hard work, simplicity, independence, courage, and justice. These nine aspects are operationalised into 45 Likert scale statement items arranged to reflect multidimensional dimensions but are measured unidimensionally. (2) The instrument developed met the validity and reliability requirements through a series of rigorous analyses. The results of content validity using Aiken's V showed that 17.7% of the items were categorised as high, 77.8% as medium, and 4.4% as low. Confirmatory Factor Analysis (CFA) confirmed the unidimensional structure of the instrument after modification, supported by optimal model fit indices (RMSEA = 0.02; CFI = 0.98; TLI = 0.98). The instrument's construct reliability was classified as very high (CR = 0.96). At the same time, IRT analysis with the Graded Response Model (GRM) showed a precise test information function (TIF) to measure students' integrity level across a wide ability range (-4.41 to 4.31). (3) Application of the instrument to 176 economics and accounting students revealed a polarization of integrity levels: 87% of respondents were at the "very high" (64%) and "high" (23%) levels, while 13% were classified as moderate to very low. Crucial findings show that honesty and responsibility are consistently the weakest aspects compared to the other seven aspects. These results highlight the urgency of focused character education interventions while also reflecting the instrument's capacity as a diagnostic tool to map students' integrity vulnerabilities.

Conflict of interests

There are no known conflicts of interest associated with this publication.

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i1.84560

REFERENCES

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings, Educational and Psychological Measurument. Journal Articles; Reports - Research; Numerical/Quantitative Data, 45(1), 131–142. https://doi.org/10.1177/0013164485451012

Alordiah, C. O., & Oji, J. (2024). Test Equating in Educational Assessment: A Comprehensive framework for Promoting Fairness, validity, and cross- cultural equity. Asian Journal of Assessment in Teaching and Learning, 14(1), 70–84. https://doi.org/10.37134/ajatel.vol14.1.7.2024

Alqahtani, M. A. (2022). Cybersecurity Awareness Based on Software and E-mail Security with Statistical Analysis. Computational Intelligence and Neuroscience, 2022, 1–12. https://doi.org/10.1155/2022/6775980

Bantam, D. J., & Nur Zhafarina, A. (2022). Academicians' Perception of the Implementation of Anti-Corruption Character Education in Higher Education. Journal An-Nafs: Kajian Penelitian Psikologi, 7(1), 88–101. https://doi.org/10.33367/psi.v7i1.2150

Catalán, H. E. N., & Gordon, D. (2020). The importance of reliability and construct validity in multidimensional poverty measurement: An illustration using the Multidimensional Poverty Index for Latin America (MPI-LA). Journal of Development Studies, 56(9), 1763–1783. https://doi.org/10.1080/00220388.2019.1663176

Cerratto Pargman, T., & McGrath, C. (2021). Mapping the Ethics of Learning Analytics in Higher Education: A Systematic Literature Review of Empirical Research. Journal of Learning Analytics, 8(2), 123–139. https://doi.org/10.18608/jla.2021.1

Chang, H.-H., Wang, C., & Zhang, S. (2021). Statistical Applications in Educational Measurement. Annual Review of Statistics and Its Application, 8(1), 439–461. https://doi.org/10.1146/annurev-statistics-042720-104044

Dewantara, J. A., Hermawan, Y., Yunus, D., Prasetiyo, W. H., Efriani, E., Arifiyanti, F., & Nurgiansah, T. H. (2021). Anti-corruption education as an effort to form students with character humanist and law-compliant. Jurnal Civics: Media Kajian Kewarganegaraan, 18(1), 70–81. https://doi.org/10.21831/jc.v18i1.38432

Dhir, A., Khan, S. J., Islam, N., Ractham, P., & Meenakshi, N. (2023). Drivers of sustainable business model innovations. An upper echelon theory perspective. Technological Forecasting and Social Change, 191, 122409. https://doi.org/10.1016/j.techfore.2023.122409

Dickerson, B. C., Atri, A., Clevenger, C., Karlawish, J., Knopman, D., Lin, P., Norman, M., Onyike, C., Sano, M., Scanland, S., & Carrillo, M. (2025). he Alzheimer's Association clinical practice guideline for the Diagnostic Evaluation, Testing, Counseling, and Disclosure of Suspected Alzheimer's Disease and Related Disorders (DETeCD-ADRD): Executive summary of recommendations for specialty care. Alzheimer's & Dementia, 21(1). https://doi.org/10.1002/alz.14337

Dipierro, A. R., & Rella, A. (2024). What lies behind perceptions of corruption? A cultural approach. Social Indicators Research, 172(2), 371–391. https://doi.org/10.1007/s11205-023-03294-4

Fierro Bósquez, M. J., Fuentes Mendoza, E. M., Olabarrieta-Landa, L., Abiuso Lillo, T., Orozco-Acosta, E., Mascialino, G., Arango-Lasprilla, J. C., & Rivera, D. (2025). Psychometric

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i1.84560

properties and normative data using item response theory approach for three neuropsychological tests in Waranka children population. Healthcare, 13(4), 423. https://doi.org/10.3390/healthcare13040423

Fitraynsyah, M. A., & Hilmiyati, F. (2024). Peran tingkat kesukaran dan daya pembeda dalam analisis butir tes: Kajian literatur untuk pendidikan menengah. Jurnal Riset Dan Evaluasi Pendidikan, 1(4), 252–262.

Germano, N. D. G., Cogo-Moreira, H., Coutinho-Lourenço, F., & Bortz, G. (2021). A new approach to measuring absolute pitch on a psychometric theory of isolated pitch perception: Is it disentangling specific groups or capturing a continuous ability? PLOS ONE, 16(2), e0247473. https://doi.org/10.1371/journal.pone.0247473

Gewers, F. L., Ferreira, G. R., Arruda, H. F. D., Silva, F. N., Comin, C. H., Amancio, D. R., & Costa, L. D. F. (2022). Principal Component Analysis: A Natural Approach to Data Exploration. ACM Computing Surveys, 54(4), 1–34. https://doi.org/10.1145/3447755

Ginanjar, D., & Purnama, W. W. (2023). Optimizing legal strategies: Combating corruption through anti-corruption education in universities. Veteran Law Review, 6(2), 122–132. https://doi.org/10.35586/velrev.v6i2.6477

Goretzko, D., Siemund, K., & Sterner, P. (2024). Evaluating Model Fit of Measurement Models in Confirmatory Factor Analysis. Educational and Psychological Measurement, 84(1), 123–144. https://doi.org/10.1177/00131644231163813

Gyamfi, A., & Acquaye, R. (2023). Parameters and models of Item esponse Theory (IRT): A review of literature. Acta Educationis Generalis, 13(3), 68–78. https://doi.org/10.2478/atd-2023-0022

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate Data Analysis. Cengage Learning.

Istiyono, E.. 2020. Pengembangan Instrumen Penilaian dan Analisis Hasil Belajar Fisika dengan Teori Tes Klasik dan Modern. (Edisi kedua). Yogyakarta: UNY Press.

Istiyono, E., Mardapi, D., & Suparno, S. (2014). Pengembangan tes kemampuan berpikir tingkat tinggi fisika (PysTHOTS) peserta didik SMA. Jurnal Penelitian Dan Evaluasi Pendidikan, 18(1), 1–12. https://doi.org/10.21831/pep.v18i1.2120

Junaidah, J., Nurbaiti, S., Riduan, R., & Amilda, A. (2022). Internalization of anti-corruption values at the University of Lampung: Integrative curriculum. AL-ISHLAH: Jurnal Pendidikan, 14(4), 5637–5644. https://doi.org/10.35445/alishlah.v14i4.2110

Lang, J. W. B., & Tay, L. (2021). The science and practice of item response theory in organizations. Annual Review of Organizational Psychology and Organizational Behavior, 8(1), 311–338. https://doi.org/10.1146/annurev-orgpsych-012420-061705

Lim, W. M. (2024). A typology of validity: Content, face, convergent, discriminant, nomological and predictive validity. Journal of Trade Science, 12(3), 155–179. https://doi.org/10.1108/JTS-03-2024-0016

Lozano-Peña, G., Sáez-Delgado, F., López-Angulo, Y., & Mella-Norambuena, J. (2021). Teachers' social–emotional competence: History, concept, models, instruments, and recommendations for educational quality. Sustainability, 13(21), 12142. https://doi.org/10.3390/su132112142

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i1.84560

Luong, R., & Flake, J. K. (2023). Measurement invariance testing using confirmatory factor analysis and alignment optimization: A tutorial for transparent analysis planning and reporting. Psychological Methods, 28(4), 905–924. https://doi.org/10.1037/met0000441

Manggaberani, A. A., & Putro, N. H. P. S. (2024). Development of character assessment instrument on English learning for middle school students. Research and Development in Education (RaDEn), 4(1), 374–389. https://doi.org/10.22219/raden.v4i1.31923

Marianti, S., Permatasari, D. P., & Rahajeng, U. W. (2021). Applying Item Response Theory model for evaluating item and test properties of academic potential test for students with disability. Jurnal Penelitian Dan Evaluasi Pendidikan, 25(1), 97–107. https://doi.org/10.21831/pep.v25i1.38808

Maulidi, A., Girindratama, M. W., Putra, A. R., Sari, R. P., & Nuswantara, D. A. (2024). Qualitatively beyond the ledger: Unravelling the interplay of organisational control, whistleblowing systems, fraud awareness, and religiosity. Cogent Social Sciences, 10(1), 2320743. https://doi.org/10.1080/23311886.2024.2320743

Nadir, N. (2024). The urgency of anti-corruption education course in universities as a long-term approach model to preventing corrupt behavior and criminal acts of corruption authors. Journal of Education Research, 5(1), 795–806. https://doi.org/10.37985/jer.v5i1.894

Okoye, J. N., Nnenna Dorothy, O., & Chigozie Ojimba, C. (2024). Corruption in the Nigerian public sector: Causes, consequences and sustainable solutions. International Journal of Advanced Multidisciplinary Research and Studies, 4(5), 1109–1117. https://doi.org/10.62225/2583049X.2024.4.5.3374

Pada, A. U. T., Mustakim, S. S., & Subali, B. (2018). Construct validity of creative thinking skills instrument for biology student teachers in the subject of human physiology. Jurnal Penelitian Dan Evaluasi Pendidikan, 22(2), 119–129. https://doi.org/10.21831/pep.v22i2.22369

Paranata, A. (2025). A systematic literature review of anti-corruption policy: A future research agenda in Indonesia. Public Organization Review, 62. https://doi.org/10.1007/s11115-025-00847-8

Pertiwi, K., & Ainsworth, S. (2021). "Democracy is the cure?": Evolving constructions of corruption in Indonesia 1994–2014. Journal of Business Ethics, 173(3), 507–523. https://doi.org/10.1007/s10551-020-04560-y

Retnawati, H. (2014). Teori Respons Butir dan Penerapannya: Untuk Peneliti, Praktisi Pengukuran dan Pengujian, Mahasiswa Pascasarjana. Nuha Medika.

Retnawati, H. (2016). Analisis Kuantitatif Instrumen Penelitian (panduan peneliti, mahasiswa, dan psikometrian). Parama Publishing.

Sanjaya, A. P., & Trifena, I. (2023). The role of education in curbing corruption: A comparison of Indonesia and Hong Kong. Integritas : Jurnal Antikorupsi, 9(2), 241–256. hhttps://doi.org/10.32697/integritas.v9i2.992

Utamirohmahsari, U. (2024). Character education building a generation with integrity and ethics. International Journal Multidisciplinary : Economics, Management, Law and Education, 1(1). https://journal-internationalmultidisciplinary.com/index.php/IJM/article/view/6

Wulandari, A., Fitriawan, R. A., Nugroho, C., Nurdiarti, R. P., Nastain, M., & Nasionalita, K. (2024). Indonesia's Women: Corruption Is a Normal Thing (Survey of Women's

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i1.84560

Perception of Corruption in Indonesia). Sage Open, 14(2), 21582440241259956. https://doi.org/10.1177/21582440241259956

Yusoff, A. M., Madihie, A., & Hutasuhut, I. J. (2023). Efikasi Kendiri sebagai moderator antara metakognitif dan resilien dalam kalangan guru bimbingan dan kaunseling di Sarawak. Malaysian Journal of Social Sciences and Humanities (MJSSH), 8(1), 1–17. https://doi.org/10.47405/mjssh.v8i1.2038

ISSN (print) 2685-7111 | ISSN (online) 2338-6061