



Confirmatory Factor Analysis in the Measurement of the TOSCA-3 Guilt Scale

Syahri Hidayahni Syam Nasution^{1*}, Debby Anggraini Daulay²
University of North Sumatera

Corresponding Author: Syahri Hidayahni Syam Nasution
shidayahni@gmail.com

ARTICLE INFO

Keywords: TOSCA-3, Analysis, Measurement, Guilt Scale

Received : 20, November

Revised : 22, December

Accepted: 31, January

©2026 Nasution, Daulay: This is an open-access article distributed under the terms of the [Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



ABSTRACT

This study aims to evaluate the measurement of guilt in adolescents using the TOSCA-3 (Test of Self-Conscious Affect 3) scale developed by Tangney and Dearing (2002). Guilt is a moral emotion that plays an important role in adolescents' psychological development, motivating them to correct behaviors and repair damaged social relationships. This research uses a quantitative approach with a non-experimental design, involving 225 adolescent respondents aged 15-19 years. Data were collected through questionnaires using an adapted TOSCA-3 scale consisting of eight aspects of guilt. The analysis results indicate that the scale has excellent reliability, with Cronbach's α of 0.867 and McDonald's ω of 0.869. However, the model fit test (CFA) revealed some fit issues, particularly with CFI and TLI values below the standard. This study suggests modifying certain items to improve the model fit. These findings indicate that the guilt measurement using TOSCA-3 is reliable, although further improvements to the instrument are needed.

INTRODUCTION

Guilt in adolescents is one of the moral emotions that plays an important role in their psychological development. During adolescence, this feeling functions as a mechanism to enhance moral awareness and support behavioral adjustment toward greater responsibility. Guilt encourages adolescents to repair damaged relationships and to align their behavior with social norms accepted by their social groups.

According to Tangney and Dearing (2002), guilt can arise from an individual's failure to meet personal or interpersonal standards and values. Tangney states that guilt, when not well managed, can become a psychological burden. She further explains that guilt focuses on behavior rather than on the self as a whole. An individual can be a good person while still engaging in bad behavior. Thus, when people feel guilty, they typically feel bad about something specific they have done, something concrete, or something they failed to do that they believe they should have done.

Tangney and Fischer (1995) explain that guilt is often triggered by failures to meet the expectations of parents or peers, which in turn motivates adolescents to correct their mistakes. Guilt functions to increase awareness of the importance of social values and to guide adolescents toward more responsible behavior in their social relationships. By experiencing guilt, adolescents are able to repair damaged relationships and adjust their behavior to better conform to accepted social norms.

To assess guilt, a number of measurement instruments have been developed and used by researchers. One widely referenced instrument is the TOSCA-3 (Test of Self-Conscious Affect 3), designed by Tangney and Dearing (2002). This scale measures adolescents' tendency toward guilt through several key aspects, including focus of evaluation, level of distress or sadness, phenomenological experience, internal operating systems, impact on the self, focus on the impact on others, counterfactual processes, and motivational features. The scale uses four response options, with higher scores indicating more intense, frequent, and impactful experiences of guilt on an individual's psychological well-being.

A study by Cokelet and Maley (2019) using the TOSCA-3 found that adolescents who are more prone to feeling guilt after making mistakes are more likely to engage in prosocial behaviors, such as helping others and repairing social relationships. In contrast, adolescents who are more prone to experiencing shame tend to feel more isolated and anxious, which can negatively affect their relationships with peers and family members.

Guilt is an important factor in motivating adolescents to make positive changes in their behavior. It often leads to more objective self-evaluation, encouraging adolescents to improve their behavior in accordance with the moral values they hold. However, instruments available to measure guilt remain limited and have not been extensively developed, particularly in Indonesia. This limitation highlights the need for the development of more comprehensive measurement tools that are better suited to the current research context.

LITERATURE REVIEW

Item Discrimination

Item diskriminasi merujuk pada sejauh mana suatu item dapat membedakan individu dengan tingkat kemampuan yang berbeda dalam konstruk yang diukur. Semakin tinggi faktor loading suatu item pada faktor laten, semakin baik item tersebut dalam membedakan individu dengan skor tinggi dan rendah pada konstruk yang diukur (Thompson, 2004). A good item discrimination value is considered to be greater than 0.30

Validity

Validity is a measure used to determine whether a test truly measures what it claims to measure (Gravetter, 2017). In CFA, construct validity is divided into two main types: convergent validity, which assesses how strongly items measuring the same construct are correlated, and discriminant validity, which assesses how weakly items measuring different constructs are correlated (Thompson, 2004; Brown, 2015).

Reliability

In addition to evaluating the validity of a measurement procedure, correlations are also used to determine reliability. A measurement procedure is considered reliable to the extent that it produces stable and consistent measurements. That is, a reliable measurement procedure will produce the same or nearly the same scores when the same individuals are measured twice under the same conditions (Gravetter, 2017). In CFA, reliability is assessed by measuring the internal consistency of the items within a scale. A reliable instrument will provide consistent results when used repeatedly under similar conditions. High factor loadings indicate good reliability of the item in measuring the latent construct. A standard for reliability is that the coefficient (e.g., Cronbach's alpha or CFA-based reliability estimates) should generally exceed 0.70 to be considered acceptable for research purposes. This threshold indicates that the measurement is sufficiently consistent and dependable (Brown, 2015).

METHODOLOGY

Population and Sampling Technique

The population of this study consists of adolescents aged 15–19 years. The study involved a sample size exceeding five times the number of items analyzed, which meets the minimum sample size recommended for obtaining valid data in factor analysis (Hair et al., 2018). The participants in this study were senior high school students, with a total of 225 respondents. The sampling method used was non-probability sampling with a convenience sampling technique.

Table 1. Guilt Scale

No.	Aspect	Item		Total Item
		<i>Favourable</i>	<i>Unfavourable</i>	
1	Evaluation Focus (EF)	1,17	2,18	4
2	Level of Distress (LD)	3,19,33	4,20,34	6
3	Phenomenological Experience (PE)	5,21	6,22	4

4	Internal Operating System (IOS)	7,23	8,24	4
5	Impact on the Self (IS)	9,25,35	10,26,36	6
6	Focus on Consequences for Others (FCO)	11,27	12,28	4
7	Counterfactual Processing (CP)	13,29	14,30	4
8	Motivational Features (MF)	15,31,37	16,32,38	6
Total		19	19	38

Data Analysis Technique

The data analysis technique involved testing the construct validity of the TOSCA-3 scale using Confirmatory Factor Analysis (CFA) and evaluating model fit. In addition, reliability testing was conducted to assess the consistency of the measurement instrument, ensuring that the results are reliable and trustworthy. Reliability analysis was performed using Cronbach’s Alpha (Azwar, 2014).

RESULTS AND DISCUSSION

This study involved 225 adolescents. All data were collected through an online questionnaire (Google Form) and analyzed using Confirmatory Factor Analysis (CFA) with Jamovi software version 2.6.44. After conducting the CFA, the results of the model fit and reliability testing were obtained as follows.

Table 2. TOSCA-3 Reliability

Scale Reliability Statistics		Description
Cronbach's α	McDonald's ω	Reliabel
0.867	0.869	

Based on Table 2, the scale reliability statistics show a Cronbach’s α value of 0.801 and a McDonald’s ω value of 0.824, indicating that the scale demonstrates very good reliability. Both values exceed the commonly accepted threshold of 0.70 as suggested by Brown (2015). These findings indicate high internal consistency of the instrument. In other words, the scale can be considered reliable for consistently measuring guilt among respondents.

Table 3. TOSCA-3 Item Discrimination

Aitem	Item-rest correlation	Aitem	Item-rest correlation
EF1	0.1681	IS2	0.5478
EF2	0.4188	IS3	0.3378
EF3	-0.0123	IS4	0.5491
FE4	0.6447	IS5	-5.080
LD1	0.3269	IS6	-0.1815
LD2	0.2092	FCO1	0.4183
LD3	0.5421	FCO2	0.1566
LD4	-0.3049	FCO3	0.4102
LD5	0.4135	FCO4	0.3724

LD6	0.2687	CP1	0.2500
PE1	-0.2437	CP2	0.2956
PE2	0.5054	CP3	0.3499
PE3	-0.0855	CP4	0.6004
PE4	0.5408	MF1	0.4859
IOS1	0.3622	MF2	0.4870
IOS2	0.2810	MF3	0.4945
IOS3	0.2047	MF4	-0.0623
IOS4	0.4967	MF5	0.4880
IS1	-0.1108	MF6	0.2458

Item discrimination analysis was conducted to ensure that each item in the scale is capable of adequately distinguishing among items measuring the same construct. Referring to the criterion that an item-rest correlation value greater than 0.30 indicates acceptable validity (Azwar, 2014), the results presented in Table 3 show that, overall, most items in the scale demonstrate adequate item-rest correlation values, with an average above 0.30. This indicates that these items contribute meaningfully to the measured construct and possess sufficient discriminative power.

Table 4. Model Fit Indices

Fit Test	Value	Data Fit	Description
χ^2	2.17	Acceptable	Acceptable Fit
SRMR	0.0973	Acceptable	Marginal Fit
RMSEA	0.077	Acceptable	Reasonable Fit
CFI	0.676	Not Acceptable	Poor Fit
TLI	0.622	Not Acceptable	Poor Fit

A model is considered to have a good fit when it meets the following criteria: Chi-square probability ($p > .05$), Root Mean Square Error of Approximation (RMSEA $< .06$), Comparative Fit Index (CFI $> .90$), Tucker-Lewis Index (TLI $> .90$), Goodness of Fit Index (GFI $> .90$), and Standardized Root Mean Square Residual (SRMR $< .08$) (Brown, 2006; Wang & Wang, 2019). Based on Table 4, the model fit results show that the chi-square value (χ^2) of 2.17 indicates that the model demonstrates an acceptable fit, suggesting that the overall model is reasonably consistent with the observed data. The SRMR value of 0.0973 indicates a marginal fit, meaning that the model shows slight deviations from the expected data. The RMSEA value of 0.077 also reflects a reasonable fit, indicating that the model is acceptable although there is still room for improvement. However, the CFI value of 0.676 and the TLI value of 0.622 both indicate poor fit, suggesting significant issues in how well the model represents the relationships among the observed variables. Overall, although some fit indices indicate acceptable results, the model requires further refinement to improve its overall fit. Therefore, the researcher proceeded to propose item-level modifications based

on the modification indices (MI) and factor loadings obtained from the CFA results, as presented below.

Table 5. Summary Modification Indices Result

Item	Factor Loadings	MI Factor Loadings	Highest MI Factor Loadings	MI Residual Covariance	Highest MI Residual Covariance
EF1	0.3668	EF, LD, IS, FCO	6.0083	EF3, EF4	EF1
EF3	0.4527	EF, LD, PE, IOS, FCO	10.5208	EF1, EF4	EF3
LD1	0.2213	-	0.5212	LD2, LD4, LD6	LD1
LD2	0.2098	LD, PE, IOS, FCO, CP	15.2070	LD1, LD3, LD6	LD2
LD4	0.2098	EF, LD, PE, IS, CP	7.0447	LD1, LD2, LD6	LD4
LD6	-0.5099	LD, IOS, IS, FCO	6.7727	LD1, LD2, LD4	LD6
PE3	-0.5742	PE, IOS, CP	4.6623	PE2, PE4	PE3
IOS1	0.2827	PE, IOS, IS, FCO	1.1863	IOS2, IOS3, IOS4	IOS1
IOS3	0.1388	IOS, IS, CP	0.1066	IOS2, IOS3	IOS3
IS1	-0.4669	IS, MF, FCO	0.0107	IOS1, IOS3	IS1
IS2	-0.2006	PE	5.0371	IS1, IS4	IS2
IS4	-0.4680	IS, MF, IOS, CP	2.9584	IS2, IS4	IS4
IS6	0.0919	LD, PE, IOS, FCO, CP, MF	19.5561	IS1, IS3	IS6
FCO2	0.1192	FCO, CP, MF	13.2691	IS1, IS4	FCO2
FCO4	0.3850	FCO, CP, MF	34.0551	IS1, IS2	FCO4
CP1	0.2362	CP, MF, IOS, FCO	0.3687	FCO1, FCO4	CP1
CP3	0.3461	CP, MF, PE	0.0609	FCO1, FCO2	CP3
MF2	0.4666	MF, CP, PE, FCO	20.6904	CP3, CP4	MF2
MF4	0.1492	EF, PE, IOS, IS, FCO, MF	26.0645	MF2, MF3, MF6	MF4
MF6	-0.1575	MF, CP, IS, PE	9.1037	CP1, CPF4	16.1973

Based on Table 5, the analysis of Modification Indices (MI) indicates the relationships among items within the same factor as well as across different factors, as reflected by MI factor loadings and MI residual covariance values. Within the Evaluation Focus (EF) dimension, items requiring modification are EF1 and EF3. In the Level of Distress (LD) dimension, items requiring modification include LD1, LD2, LD4, and LD6. For the Phenomenological Experience (PF) dimension, items requiring attention are PF3 and PF2. In the Internal Operating System (IOA) dimension, items requiring modification are

IOS1 and IOS3. In the Impact on the Self (IS) dimension, items requiring modification include IS1, IS2, IS4, and IS6. For the Focus on Consequences for Others (FCO) dimension, items requiring modification are FCO2 and COI4. In the Counterfactual Processing (CP) dimension, items requiring modification are CP1 and CP3. Finally, within the Motivational Features (MF) dimension, items requiring attention are MF2, MF4, and MF6. Overall, these items demonstrate potential for improving inter-item relationships within their respective dimensions, which is expected to enhance the overall model fit.

The study on the TOSCA-3 guilt scale demonstrated that the scale has high reliability, with Cronbach's α and McDonald's ω values both exceeding the recommended threshold of 0.70, indicating good internal consistency for measuring guilt in adolescents. Despite its reliability, the Confirmatory Factor Analysis (CFA) revealed several issues with the model fit. While the chi-square value indicated an acceptable fit, other index, such as the Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI), were below the acceptable standards, suggesting that the model did not fully capture the relationships between the observed variables. This finding highlights the importance of further model modifications to improve fit, such as item-level adjustments based on modification indices and factor loadings, which could lead to a better representation of the latent factors.

Furthermore, the study emphasized the need for the development of more suitable measurement tools, particularly in the Indonesian context, where guilt measurement instruments are still limited. Modifying items, especially those with low item-rest correlations, can improve the scale's effectiveness in accurately capturing the dimensions of guilt. The results also underline the significant role of guilt in adolescents' moral development, as it motivates positive behavioral changes, self-evaluation, and the repair of social relationships. Future studies should consider revising the problematic items and validating the instrument in different cultural contexts to enhance its applicability and reliability across diverse populations.

CONCLUSIONS AND RECOMMENDATIONS

Based on the research findings, it can be concluded that the TOSCA-3 scale demonstrates excellent reliability in measuring guilt in adolescents, with Cronbach's α and McDonald's ω values exceeding the recommended threshold. However, the Confirmatory Factor Analysis (CFA) revealed issues with model fit, particularly with CFI and TLI values below the accepted standards, indicating that the model needs modification. Item modifications based on modification indices and factor loadings could improve the overall model fit. This study also highlights the importance of developing more suitable measurement tools, particularly within the Indonesian context, and validating the instrument across different cultures to enhance its applicability and reliability across broader populations.

FURTHER STUDY

Future studies may consider removing items that are deemed irrelevant, subject to approval from Tangney, or conducting further testing after implementing model modifications.

ACKNOWLEDGMENT

The author would like to express sincere gratitude to all research participants who generously took the time to complete the questionnaire and share their experiences. Appreciation is also extended to the academic supervisor and colleagues in the Professional Psychology Master's Program, Faculty of Psychology, Universitas Sumatera Utara, for their guidance, valuable feedback, and support throughout the research process. In addition, the author acknowledges those who assisted in the distribution of the questionnaire and data processing. This study would not have been completed without the valuable contributions of all involved parties.

REFERENCES

- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.
- Cokelet, B., & Maley, C. J. (Eds.). (2019). *The moral psychology of guilt*. Rowman & Littlefield.
- Creswell, J.W., & Creswell J.D. (2018). *Research Design Qualitative, Quantitative, and Mixed*
- Hair, J. F., et al. (2018). *Multivariate data analysis (8th Edition)*. United Kingdom: Cengage Learning. *Methods Approaches*
- Tangney, J. P., & Fischer, K. (1995). *SELF-CONSCIOUS EMOTIONS (Vol. 1, Issue 1)*.
- Tangney, J. P., & Ronda L. Dearing. (2002). *Shame And Guilt*. In P. Salovey (Ed.), *The Guilford Press (Series Edi)*.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. American Psychological Association.
- Wang, J., & Wang, X. (2019). *Structural equation modelling: Application using Mplus 2nd Ed*. John Wiley & Sons Ltd.