The 4th International Conference on Education Innovation and Social Science, August 2025
ISSN (Online): 2961-9602

CONSTRUCTING AND VALIDATING A CONTEXTUAL GRIT SCALE FOR UNIVERSITY STUDENTS IN INDONESIA: A PSYCHOMETRIC STUDY

Muhamad Syamsul Hayat[1], Rika Dwi Agustiningsih[2], Novita Sari[3]

[1,2,3]Department of Psychology, Faculty of Social and Humanities, Universitas Muhammadiyah Bandung, Indonesia
[1]muhamadsyamsul3798@gmail.com

Keyword

Grit, scale development, reliability, validity, psychometrics.

Abstract

This study aims to develop and evaluate the psychometric properties of a grit measurement instrument for university students in Indonesia, with a specific focus on the Bandung region. The theoretical framework conceptualized grit as comprising two core dimensions: perseverance of effort and consistency of interest. Thirty-two items were initially developed based on seven behavioral indicators and evaluated through expert judgment, language clarity tests, and item discrimination analysis. After a series of empirical tests involving 100 students for the tryout, 200 for the field test, and 40 for the test–retest, 14 items were retained based on reliability and validity criteria. Confirmatory Factor Analysis (CFA) supported the two-dimensional model with acceptable goodness-of-fit indices (RMSEA = 0.032, CFI = 0.977, TLI = 0.964, SRMR = 0.044). Internal consistency reliability was satisfactory (Cronbach's α = 0.790), and test–retest results demonstrated high temporal stability (r = 0.953, $p < .001$). These findings indicate that the developed grit scale is a valid and reliable instrument for assessing grit among Indonesian university students and can be applied in educational psychology research and interventions.

INTRODUCTION

Students undergo a pivotal stage of intellectual and personal development as they mature. Within educational psychology, particularly from a psychometric standpoint, students are conceptualised not only as recipients of academic instruction but also as individuals whose cognitive, affective, and behavioural attributes can be systematically measured and evaluated. In the Indonesian education system, the formal educational trajectory typically comprises nine years of basic education, three years of secondary education, and approximately four years of higher education. Throughout this progression, students are expected to assimilate academic knowledge, apply it within relevant contexts, and complete their educational journey effectively and efficiently. Higher education institutions serve as centres for academic learning and platforms for cultivating personal discipline, responsibility, and readiness for professional integration.

Accurate assessment of students' psychological characteristics through valid and reliable measurement tools is thus fundamental to enhancing educational outcomes and fostering holistic student development. Nonetheless, the academic journey at the tertiary level is often fraught with challenges. Many students encounter difficulties such as procrastination, diminished motivation,

and an inability to sustain the long-term enthusiasm necessary to achieve academic goals. Addressing these challenges requires the psychological trait known as *grit*.

As Duckworth (2007, 2016) introduced, grit refers to the perseverance of effort and consistency of interest over extended periods in pursuit of long-term objectives. Duckworth initially developed the Grit-O scale, which was subsequently refined into the more concise Grit-S version, demonstrating robust psychometric properties across diverse cultural contexts. However, applying Grit-S within the Indonesian cultural and linguistic context—particularly among university students—presents several limitations, including potential linguistic ambiguities and cultural incongruence. These issues underscore the need for a culturally relevant and contextually appropriate measure of grit for Indonesian students.

Efforts to address this gap have been initiated. For example, Indraswari (2020) developed a grit scale grounded in local cultural nuances, which exhibited relatively high reliability. However, methodological limitations, such as insufficient heterogeneity in the sample, continue to challenge generalizability. Moreover, while several studies in Indonesia have adapted Duckworth's grit instruments, many have not adhered rigorously to the International Test Commission (ITC) guidelines for test adaptation—such as employing forward-backwards translation, ensuring cross-cultural validity, and maintaining comprehensive documentation.

Accordingly, developing a valid, reliable, and psychometrically sound grit scale has become a pressing priority in educational psychology research in Indonesia. Such an instrument is essential for evaluation and provides a foundation for intervention strategies to enhance student persistence, learning outcomes, and academic achievement. This study seeks to develop and validate a grit scale suited to the Indonesian student population—particularly in the Bandung region—by incorporating established principles of psychometric development and ensuring adequate representation of diverse groups. The resulting scale is a practical tool for academic research and educational applications.

METHOD

This study employed a quantitative research design with a psychometric analysis approach, utilising quota sampling as the non-probability sampling method. Participants were selected based on predefined criteria, sample size requirements, and research objectives. The inclusion criterion specified that participants must be actively enrolled university students in Bandung. Data were collected online via a Google Form, which the researcher distributed through social media platforms, including WhatsApp and Instagram. The study included 100 participants in the preliminary try-out, 200 in the main data collection, and 40 in the retest, all of whom met the inclusion criteria.

A five-point Likert scale comprised response options ranging from *Strongly Agree* to *Strongly Disagree*. The development of the grit scale began with an extensive literature review to conceptualise grit, adopting Duckworth et al (2007) widely recognised definition. Based on this framework, the construct was operationalised as two dimensions and seven indicators, yielding 32 items, both positively and negatively worded.

The initial item pool was structured according to a theoretical blueprint and evaluated by three subject-matter experts using expert judgment to assess content validity. A language clarity test was also conducted to ensure item comprehensibility. A tryout with 100 participants was administered to determine item discrimination and reliability, with a minimum acceptable item-total correlation of 0.30 and a Cronbach's alpha threshold of 0.790. Data analyses were conducted using SPSS and JASP, including Confirmatory Factor Analysis (CFA) to examine the scale's construct validity.

A field test was then conducted with 200 students. Items with discrimination indices below 0.30, negative item-total correlations, or incomplete responses were excluded. Of the original 32 items, 14 met the psychometric criteria and demonstrated model fit based on CFA results. The finalised scale comprised eight items representing perseverance of effort and six

items reflecting consistency of interest. A test–retest procedure conducted with 40 students over two weeks indicated stable and statistically significant reliability over time.

Table 1.
Blue Print

Variable	Aspects	Indicators	Favorable	Unfavorable
Grit	Kegigihan dalam Berusaha	Menyelesaikan segala sesuatu yang telah dimulai	2	2
		Berusaha keras dalam menghadapi tantangan	2	2
		Berusaha keras dalam menghadapi hambatan	4	1
		Berusaha mencapai tujuan jangka panjang	3	2
	Konsisten Pada Minat	Mengerjakan hal yang sama sesuai minat secara terus menerus	2	3
		Memiliki tujuan jangka panjang	3	2
		Mempertahankan minat dalam jangka panjang	2	2

RESULTS

In the data collection process, respondent distribution revealed that gender representation was balanced during the try-out phase, with 50 male and 50 female participants. However, female respondents outnumbered male respondents in both the field test and test–retest phases. Specifically, the field test comprised 145 females and 55 males, while the test–retest involved 28 females and 12 males. This disproportionate participation suggests a predominance of female involvement, which may be attributed to greater interest in psychological research or increased accessibility and willingness to complete online questionnaires.

Regarding age distribution, most respondents in the tryout and field test phases were 20 years old, accounting for 27 individuals (27% and 27.5%, respectively). In contrast, during the test–retest phase, the age group represented the most was 22 years, with 18 respondents (45%). These findings indicate that students aged 20 and 22 were the most actively engaged in the study. Detailed demographic distributions are presented in the table below.

Table 2.
Demographic Data Based on Gender

	Try Out		*Field Test*		*Test Retest*	
	F	%	F	%	F	%
Male	50	50%	55	27,5%	12	30%
Female	50	50%	145	72,5%	28	70%

Based on age, the following age ranges were obtained.

M

Table 3.
Demographic Data Based on Age

	Try Out		*Field Test*		*Test Retest*	
	F	%	F	%	F	%
18	3	3%	7	3,5%	-	-
19	3	3%	37	18,5%	-	-
20	27	27%	55	27,5%	5	12,5%
21	26	26%	48	24%	12	30%
22	22	22%	32	16%	18	45%
23	12	12%	13	6,5%	5	12,5%
24	4	4%	4	2%	-	-
25	3	3%	4	2%	-	-

Psychometric Property Analysis

In the initial phase of instrument development, content validity was assessed through expert judgment of the 32 items developed. Three subject matter experts independently evaluated the relevance of each item using a 5-point Likert scale, ranging from 1 (not relevant) to 5 (highly appropriate). The assessments were analysed using Aiken's V formula to quantify the degree of agreement among the experts regarding item relevance. The overall Aiken's V coefficient was calculated at 0.92, indicating a high level of content validity across all items. These findings suggest that the expert panel deemed the items highly relevant. The detailed results of the expert judgment analysis using Aiken's V are presented in the following table.

Table 4.
Expert Judgment Results

Item	Assessor			s1	s2	s3	∑s	n(c-1)	V	Description
	I	II	III							
Item_01	5	4	5	4	3	4	11	12	0,916	Valid
Item_02	4	5	5	3	4	4	11	12	0,916	Valid
Item_03	5	4	5	4	3	4	11	12	0,916	Valid
Item_04	5	5	5	4	4	4	12	12	1	Valid
Item_05	5	5	4	4	4	3	11	12	0,916	Valid
Item_06	5	4	5	4	3	4	11	12	0,916	Valid
Item_07	4	5	5	3	4	4	11	12	0,916	Valid
Item_08	4	5	5	3	4	4	11	12	0,916	Valid
Item_09	5	5	5	4	4	4	12	12	1	Valid
Item_10	4	5	5	3	4	4	10	12	0,916	Valid
Item_11	5	5	5	4	4	4	12	12	1	Valid
Item_12	5	5	5	4	4	4	12	12	1	Valid
Item_13	4	5	5	3	4	4	11	12	0,916	Valid
Item_14	4	5	5	3	4	4	11	12	0,916	Valid
Item_15	4	5	5	3	4	4	11	12	0,916	Valid
Item_16	4	5	5	3	4	4	11	12	0,916	Valid
Item_17	4	5	5	3	4	4	11	12	0,916	Valid
Item_18	4	5	5	3	4	4	11	12	0,916	Valid
Item_19	5	5	5	4	4	4	12	12	1	Valid
Item_20	5	5	5	4	4	4	12	12	1	Valid

Item_21	4	5	5	3	4	4	11	12	0,916	Valid
Item_22	5	5	5	4	4	4	12	12	1	Valid
Item_23	4	5	5	3	4	4	11	12	0,916	Valid
Item_24	4	5	5	3	4	4	11	12	0,916	Valid
Item_25	5	4	5	4	3	4	11	12	0,916	Valid
Item_26	5	4	5	4	3	4	11	12	0,916	Valid
Item_27	5	4	5	4	3	4	11	12	0,916	Valid
Item_28	5	5	5	4	4	4	12	12	1	Valid
Item_29	4	5	5	3	4	4	11	12	0,916	Valid
Item_30	5	5	5	4	4	4	12	12	1	Valid
Item_31	4	5	5	3	4	4	11	12	0,916	Valid
Item_32	4	5	5	3	4	4	11	12	0,916	Valid

Following the expert judgment phase, a language clarity test was conducted to ensure that the items in the student grit measurement instrument were understandable to the target population. Two students were randomly selected to review and evaluate the clarity and comprehensibility of the items, ensuring alignment with prospective respondents' cognitive and linguistic levels.

Subsequently, a preliminary tryout was administered to 100 students to assess the scale's reliability and item discrimination. The results indicated that all items demonstrated satisfactory discriminating power, with item-total correlation coefficients predominantly exceeding the threshold of 0.30. Furthermore, internal consistency analysis using Cronbach's alpha yielded a reliability coefficient of 0.868. According to Azwar (2014), this value reflects a high level of reliability and strong internal consistency of the instrument. The reliability analysis results using Cronbach's alpha are presented in the following table.

Table 5.
Reliability Score During Try-outs

Cronbach's Alpha	N of Items
.868	32

The following table shows the item discrimination power test:

Table 6.
Item Discrimination Test Scores During Try-outs

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
VAR00001	73.47	132.595	.488	.862
VAR00002	72.56	133.663	.440	.863
VAR00003	73.54	132.271	.346	.866
VAR00004	74.13	135.569	.305	.866
VAR00005	72.90	132.556	.462	.862
VAR00006	72.72	131.476	.519	.861
VAR00007	72.45	134.917	.424	.864
VAR00008	74.30	135.000	.291	.867
VAR00009	72.86	134.162	.341	.865
VAR00010	74.32	134.927	.340	.865
VAR00011	72.45	132.957	.530	.862

VAR00012	74.10	127.222	.644	.857
VAR00013	73.89	136.402	.275	.867
VAR00014	73.02	134.383	.312	.866
VAR00015	73.91	129.133	.583	.859
VAR00016	73.92	133.630	.427	.863
VAR00017	73.73	131.492	.574	.860
VAR00018	72.87	135.771	.319	.866
VAR00019	74.33	141.961	-.049	.875
VAR00020	74.23	146.603	-.228	.883
VAR00021	73.22	134.133	.304	.866
VAR00022	73.10	131.444	.456	.862
VAR00023	72.73	129.856	.614	.859
VAR00024	72.67	130.506	.659	.859
VAR00025	74.24	131.417	.507	.861
VAR00026	73.19	125.529	.553	.859
VAR00027	74.13	133.670	.364	.865
VAR00028	73.33	132.789	.560	.861
VAR00029	74.16	137.408	.204	.868
VAR00030	73.63	133.124	.452	.863
VAR00031	72.70	133.404	.421	.863
VAR00032	74.24	134.548	.344	.865

The subsequent phase involved conducting a field test to evaluate the measurement instrument's feasibility and applicability across a broader and more diverse population. An empirical study was conducted with 200 actively enrolled university students in Bandung. This phase aimed to assess the instrument's psychometric properties in a real-world educational context and ensure its relevance and functionality within the target population.

Version 32 Items

The results of the item discrimination analysis indicated that most items had item-total correlation coefficients of 0.30 or greater, suggesting that the items effectively differentiated between respondents with varying levels of grit. Furthermore, internal consistency analysis yielded a Cronbach's alpha of 0.871, indicating high reliability and supporting the psychometric soundness of all items in the instrument. The detailed results of the item discrimination analysis are presented in the following table:

Table 7.
Testing the Power of 32 Items During Field Tests

Item	Item-test Correlation	Mean	SD
Item 1	0.453	2.790	1.226
Item 2	0.397	3.030	1.056
Item 3	0.237	2.515	0.972
Item 4	0.305	2.305	0.983
Item 5	0.414	2.830	0.803
Item 6	0.551	2.675	1.194
Item 7	0.428	2.810	1.209
Item 8	0.043	1.905	0.917
Item 9	0.440	3.015	0.824
Item 10	0.326	1.620	0.787

Item 11	0.477	2.790	1.189
Item 12	0.528	2.000	0.891
Item 13	0.322	2.565	1.030
Item 14	0.340	2.750	0.831
Item 15	0.436	2.225	0.622
Item 16	0.393	2.710	0.965
Item 17	0.354	2.175	0.698
Item 18	0.435	2.925	0.750
Item 19	0.062	1.805	0.787
Item 20	0.054	1.305	0.852
Item 21	0.534	2.865	0.884
Item 22	0.437	2.785	0.844
Item 23	0.563	3.065	0.757
Item 24	0.442	3.035	1.029
Item 25	0.430	1.965	0.726
Item 26	0.523	2.775	1.209
Item 27	0.450	1.890	0.861
Item 28	0.579	2.540	0.656
Item 29	0.300	2.050	0.878
Item 30	0.384	2.080	0.853
Item 31	0.483	2.705	1.097
Item 32	0.464	2.075	0.992

The following table shows reliability using Cronbach's alpha formula:

Table 8.
Reliability Value of 32 Items During Field Test

Estimate	Cronbach's α
Point estimate	0.871
95% CI lower bound	0.844
95% CI upper bound	0.894

Following the item discrimination and reliability analyses of the 32 researcher-developed items, a validity assessment was conducted using Confirmatory Factor Analysis (CFA). CFA was employed to ensure that the items measuring the grit construct among students were distinct and collectively formed a statistically acceptable model. The analysis was conducted in JASP to assess the extent to which the selected items measured the latent construct of student grit.

The initial model, based on 32 items that met the criteria for discriminant validity, was revised in one round to achieve an acceptable level of model fit. The CFA was conducted on two core dimensions of grit: perseverance of effort (represented by items 1, 2, 3, 5, 8, 11, 13, 15, 17, 18, 22, 23, 24, 26, 27, 30, 31, 32) and consistency of interest (represented by items 4, 6, 7, 9, 10, 12, 14, 16, 19, 20, 21, 25, 28, 29). This analysis aimed to confirm the multidimensional nature of the grit construct in university students.

The CFA results, computed using JASP, indicated that the revised model met the required fit indices for a second-order confirmatory factor model. The results of the Goodness-of-Fit analysis are presented in the following table.

Table 9.
Second Order Confirmation Factor Analysis Value

Goodness of fit index	Cut of value	Hasil model	Description
RMSEA	< 0,05	0,035	Fit
SRMR	< 0,08	0,068	Fit
CFI	> 0,9	0,947	Fit
TLI	> 0,9	0,932	Fit

The table above shows that the model demonstrated an acceptable level of fit, with goodness-of-fit indices meeting established thresholds for adequacy. According to Brown (2015), a good model fit is typically indicated by a Root Mean Square Error of Approximation (RMSEA) value is < 0.05, a Standardized Root Mean Square Residual (SRMR) value is > 0.08, a Comparative Fit Index (CFI) is > 0.90 or approaching 1.00, and a Tucker-Lewis Index (TLI) is > 0.90 or nearing 1.00.

Table 10.
Blueprint 32 Items

Aspects	Indicators	Items	Keterangan	No
Kegigihan Dalam Berusaha	Menyelesaikan segala sesuatu yang telah dimulai	Saya memiliki komitmen yang tinggi dalam menyelesaikan setiap tugas	Favorable	23
		Saya selalu berusaha menyelesaikan seluruh tugas mata kuliah tepat waktu	Favorable	1
		Saya lebih memilih bermain/berleha-leha daripada menyelesaikan tugas	Unfavorable	15
		Saya belum memaksimalkan seluruh usaha untuk menyelesaikan tugas	Unfavorable	32
	Berusaha keras dalam menghadapi tantangan	Saya selalu berusaha mencari referensi lain dalam menyelesaikan tugas agar mendapatkan hasil yang maksimal	Favorable	2
		Ketika mendapatkan	Unfavorable	8

	tugas dengan tenggat waktu yang singkat, saya cenderung belum memaksimalkan usaha saya dalam mengerjakannya		
	Walaupun saya sedang sakit, saya akan berusaha menyelesaikan tugas sesuai tenggat waktu	Favorable	30
	Ketika dihadapkan dengan tugas yang sulit, saya menyelesaikan tugas tidak tepat waktu	Unfavorable	17
Berusaha keras dalam menghadapi hambatan	Meskipun kesulitan dalam mencari referensi yang relevan dengan tugas perkuliahan, saya akan terus berusaha untuk mencarinya	Favorable	24
	Saya tidak tertarik mempelajari kembali materi perkuliahan yang sulit saya pahami	Unfavorable	3
	Saya akan berusaha mempelajari kembali materi perkuliahan yang sulit saya pahami	Favorable	22
	Saya berusaha mendalami materi perkuliahan yang sulit, meskipun butuh waktu yang lama	Favorable	5
	Saya terus mempelajari hal baru untuk mendapatkan Solusi dari tugas	Favorable	18

		yang sulit		
	Berusaha mencapai tujuan jangka panjang	Walaupun harus melalui proses yang panjang, saya tetap berusaha mempertahankan cita-cita saya	Favorble	26
		Saya tetap berusaha untuk menggapai cita-cita, meskipun mengalami kendala dalam proses menggapainya	Favorable	11
		Saya terus berusaha mencari tahu hal-hal yang berkaitan dengan cita-cita saya, agar cita-cita saya dapat tercapai	Favorable	31
		Ketika menghadapi kendala, saya enggan untuk tetap berusaha menggapai cita-cita	Unfavorable	27
		Ketika saya tidak mencapai target yang sudah saya tetapkan, saya merasa putus asa untuk menggapai cita-cita saya	Unfavorable	13
Konsisten pada Minat	Mengerjakan hal yang sama sesuai minat secara terus menerus	Saya mampu memprioritaskan waktu dan energi saya pada kegiatan yang sejalan dengan minat saya	Favorable	9
		Saya senang melakukan kegiatan yang berkaitan dengan minat saya dalam waktu yang lama	Favorable	7
		Saya cenderung mudah bosan ketika harus	Unfavorable	29

	mengerjakan hal yang sama sesuai minat saya dalam waktu yang lama		
	Saya kesulitan dalam memprioritaskan waktu dan energi saya pada kegiatan yang sejalan dengan minat saya	Unfavorable	4
	Kegiatan baru yang tidak sejalan dengan minat saya mengalihkan saya dari minat yang sedang saya tekuni	Unfavorable	20
Memiliki tujuan jangka panjang	Saya sedang berusaha untuk mewujudkan cita-cita saya	Favorable	28
	Saya mengetahui apa yang harus saya usahakan untuk dapat menggapai cita-cita saya	Favorable	6
	Saya merasa bingung dengan apa yang ingin saya capai di masa depan	Unfavorable	25
	Saya belum memiliki perencanaan untuk masa depan	Unfavorable	12
	Saya memiliki perencanaan yang matang untuk menggapai cita-cita saya	Favorable	21
Mempertahankan minat dalam jangka panjang	Saya selalu berusaha mengatasi kebosanan dalam proses mempertahankan minat yang saya tekuni	Favorable	14
	Jurusan yang saya	Favorable	19

	pilih sekarang adalah jurusan yang saya minati dari dulu		
	Minat saya seringkali berubah-ubah dari tahun ke tahun	Unfavorable	10
	Saya mengubah cita-cita saya, ketika proses-proses untuk menggapainya dirasa sulit	Unfavorable	16

Version 14 Items

The discriminant validity analysis indicated that most items had item-total correlation coefficients of 0.30 or greater, suggesting that the items effectively differentiated among respondents by grit level. Furthermore, the internal consistency reliability analysis yielded a Cronbach's alpha value of 0.790, which exceeds the commonly accepted threshold, indicating that the items are reliable and that the student grit scale possesses high internal consistency. The detailed results of the item discrimination analysis are presented in the following table.

Table 11.
Differential Power Test Value of 14 Items During Field Test

Item	Item-rest correlation	Mean	SD
Item 32	0.350	2.075	0.992
Item 2	0.357	3.030	1.056
Item 30	0.340	2.080	0.853
Item 22	0.451	2.785	0.844
Item 26	0.524	2.775	1.209
Item 31	0.500	2.705	1.097
Item 9	0.405	3.015	0.824
Item 7	0.406	2.810	1.209
Item 6	0.530	2.675	1.194
Item 14	0.328	2.750	0.831
Item 16	0.318	2.710	0.965
Item 15	0.394	2.225	0.622
Item 5	0.458	2.830	0.803
Item 25	0.370	1.965	0.726

The following table shows reliability using Cronbach's alpha formula:

Table 12.
Reliability Value of 14 Items

Estimate	Cronbach's α
Point estimate	0.790
95% CI lower bound	0.744
95% CI upper bound	0.829

Following the item discrimination and reliability analyses of the initial 32 items, the researcher evaluated construct validity using Confirmatory Factor Analysis (CFA). The CFA was conducted to establish a statistically acceptable, well-fitting measurement model based on the 14 items that demonstrated sufficient discriminative power. The analysis was performed in JASP to assess the extent to which the selected items accurately measured the construct of student grit.

One model revision was performed to achieve model fit. The CFA focused on two core dimensions of grit: perseverance of effort, represented by items 2, 5, 15, 22, 26, 30, 31, and 32, and consistency of interest, represented by items 6, 7, 9, 14, 16, and 25. This analysis aimed to confirm that these two dimensions adequately reflect the multidimensional nature of the grit construct among students.

Subsequent CFA analyses in JASP indicated that the model met the established goodness-of-fit threshold for a second-order confirmatory factor structure. The results of the goodness-of-fit analysis are presented in the following table.

Table 13.
Second Order Confirmatory Factor Analysis Value 14 Items During Field Test

Goodness of fit index	Cut of value	Model Output	Description
RMSEA	< 0,05	0,032	Fit
SRMR	< 0,08	0,044	Fit
CFI	> 0,9	0,977	Fit
TLI	> 0,9	0,964	Fit

As shown in the table above, the model fit indices indicated satisfactory model fit, with goodness-of-fit values within recommended thresholds. According to Brown (2015) an RMSEA value of < 0.05 indicates a good fit, an SRMR value of > 0.08 is desirable, a CFI value of > 0.90 or approaching 1.00 reflects an excellent fit, and a TLI value of > 0.90 similarly denotes a strong model fit.

Subsequently, a test–retest procedure was conducted to assess the temporal stability of the student grit measurement instrument. This procedure involved a sample of 40 students from Bandung, with data collected at two time points separated by a two-week interval. The results demonstrated a strong, statistically significant correlation between the two administrations, indicating that the instrument has high test–retest reliability. The test–retest correlation results and the corresponding scatter plot of score distributions are presented below.

Table 14.
Pearson's Correlation Value 14 Items

Variable		X	Y
1. X	Pearson's r	—	
	p-value	—	
2. Y	Pearson's r	0.953	—
	p-value	< .001	—

The Pearson correlation coefficient (r = 0.953) indicates a strong, positive relationship between the first and second test scores, indicating high temporal stability in students' grit scores

over the two-week interval. A p-value < 0.001 further supports the statistical significance of this relationship. These findings collectively provide strong evidence for the test–retest reliability of the grit measurement instrument.

Table 15.
Blueprint 14 Items

Aspects	Indicators	Items	Keterangan	No
Kegigihan Dalam Berusaha	Menyelesaikan segala sesuatu yang telah dimulai	Saya lebih memilih bermain/berleha-leha daripada menyelesaikan tugas	Unfavorable	15
		Saya belum memaksimalkan seluruh usaha untuk menyelesaikan tugas	Unfavorable	32
	Berusaha keras dalam menghadapi tantangan	Saya selalu berusaha mencari referensi lain dalam menyelesaikan tugas agar mendapatkan hasil yang maksimal	Favorable	2
		Walaupun saya sedang sakit, saya akan berusaha menyelesaikan tugas sesuai tenggat waktu	Favorable	30
	Berusaha keras dalam menghadapi hambatan	Saya akan berusaha mempelajari kembali materi perkuliahan yang sulit saya pahami	Favorable	22
		Saya berusaha mendalami materi perkuliahan yang sulit, meskipun butuh waktu yang lama	Favorable	5
	Berusaha mencapai tujuan jangka panjang	Walaupun harus melalui proses yang panjang, saya tetap berusaha mempertahankan cita-cita saya	Favorble	26
		Saya terus berusaha mencari tahu hal-hal yang berkaitan dengan	Favorable	31

		cita-cita saya, agar cita-cita saya dapat tercapai		
Konsisten pada Minat	Mengerjakan hal yang sama sesuai minat secara terus menerus	Saya mampu memprioritaskan waktu dan energi saya pada kegiatan yang sejalan dengan minat saya	Favorable	9
		Saya senang melakukan kegiatan yang berkaitan dengan minat saya dalam waktu yang lama	Favorable	7
	Memiliki tujuan jangka panjang	Saya mengetahui apa yang harus saya usahakan untuk dapat menggapai cita-cita saya	Favorable	6
		Saya merasa bingung dengan apa yang ingin saya capai di masa depan	Unfavorable	25
	Mempertahankan minat dalam jangka panjang	Saya selalu berusaha mengatasi kebosanan dalam proses mempertahankan minat yang saya tekuni	Favorable	14
		Saya mengubah cita-cita saya, ketika proses-proses untuk menggapainya dirasa sulit	Unfavorable	16

DISCUSSION

The development of the grit measurement scale began with an extensive review of the literature to conceptualise grit, with particular reference to Duckworth et al (2007) theoretical framework, which defines grit as the combination of perseverance of effort and consistency of long-term interest. Guided by this theory, the researchers identified two core dimensions and operationalised them into seven behavioural indicators that served as the foundation for item development.

Based on these indicators, 32 items—comprising both favourable and unfavourable statements—were constructed, with a 5-point Likert scale used as the response format. To assess the scale's content validity, three expert psychologists independently evaluated the items for relevance and clarity. A language clarity test was conducted to ensure the intended respondents

understood the items.

An initial trial of the scale was conducted with 100 university students to evaluate item discrimination and internal consistency. Items that failed to meet the psychometric criteria were excluded from subsequent analyses. A larger field test was then conducted with a sample of 200 active students in Bandung to examine the scale's construct validity and reliability, employing both SPSS and JASP software for statistical analysis. Items were eliminated if they demonstrated poor discrimination (item-total correlation < 0.300), negative loading values, or excessive missing responses.

Eighteen of the original 32 items were removed, resulting in a final scale of 14 items. Confirmatory Factor Analysis (CFA) supported the adequacy of the revised model, demonstrating good model fit and confirming that the items validly and reliably represented the grit construct—the final instrument comprised eight items measuring perseverance of effort and six items measuring consistency of interest.

To assess the temporal stability of the scale, a test–retest procedure was conducted with a subsample of 40 students over a two-week interval. The Pearson correlation coefficient was $r = 0.953$ ($p < .001$), indicating high consistency and strong test–retest reliability. Although the final number of items (14) was slightly lower than in some prior studies (e.g., 15 items), the scale demonstrated robust validity and reliability, supported by a larger and more diverse sample.

CONCLUSION

Based on the empirical findings from developing a grit measurement instrument for university students in Bandung, the scale comprises two primary dimensions: persistence in effort and consistency in interest, operationalised through seven behavioural indicators. The instrument is available in two formats: a short version comprising 14 items and an extended version comprising 32 items. The reliability analysis yielded Cronbach's Alpha values that exceeded the commonly accepted threshold, indicating that all items demonstrate adequate internal consistency. Furthermore, confirmatory factor analysis (CFA) provided evidence of a good model fit, supporting the hypothesised two-dimensional structure of grit and confirming that the items appropriately reflect the underlying latent constructs. These psychometric results suggest that the grit measurement instrument meets established standards of reliability and validity and may therefore be considered a robust and appropriate tool for assessing grit among university students.

ACKNOWLEDGEMENT

The authors sincerely express their gratitude to the teachers who participated in this research. Deep appreciation is also extended to colleagues and academic advisors for their constructive feedback and guidance. The authors are also grateful for the institution's support and the provision of resources that greatly contributed to the successful completion of this study.

REFERENCES

Azwar, S. (2014). *Reliabilitas dan Validitas* (4th ed.). Pustaka Pelajar.

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research* (2nd ed.). The Guilford Press.

Duckworth, A. L. (2016). *Grit: Kekuatan Passion dan Kegigihan*. Gramedia Pustaka Utama.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and Passion for Long-Term Goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087

Indraswari, C. (2020). *Penyusunan Dan Pengembangan Alat Ukur Pada Skala Pendek Grit*. 46–58.