



# DEVELOPMENT AND PSYCHOMETRIC VALIDATION OF A SELF-REPORT QUESTIONNAIRE MEASURING THE NEGATIVE IMPACT OF SOCIAL MEDIA USE ON SELF-ESTEEM AND IMPOSTER AMONG UZBEK YOUTH

Nilufar Yuldasheva

PhD, Associate Professor Department of Pedagogical  
Education and Psychology Renaissance Education University  
[e-mail: nilufary660@gmail.com]

## Abstract

**Background.** Existing self-esteem and social-comparison instruments were developed largely within Western cultural and linguistic contexts and do not fully capture the experience of social media use among Uzbek youth. This study reports the development and full psychometric evaluation of an original Uzbek-language self-report questionnaire measuring the negative impact of social media use on self-esteem.

**Methods.** The instrument was constructed over seven stages following established psychometric methodology. Content validity was evaluated by seven experts using the Content Validity Ratio (CVR) and the Item Content Validity Index (I-CVI). The main psychometric study was conducted with a stratified sample of 1,605 respondents aged 17–35 years. Internal consistency (Cronbach's alpha), exploratory factor analysis (EFA), confirmatory factor analysis (CFA), test–retest reliability ( $n = 60$ , 4-week interval), split-half reliability, and convergent and discriminant validity were assessed.

**Results.** The final 18-item scale demonstrated excellent content validity (S-CVI/Ave = .946). EFA produced a clear three-factor structure (KMO = .944; total variance explained = 55.18%), confirmed by CFA ( $\chi^2/df = 2.87$ ; CFI = .953; TLI = .944; RMSEA = .058; SRMR = .047). Internal consistency was high (total  $\alpha = .899$ ; subscale  $\alpha = .783$ –.864). Test–retest reliability ( $r = .927$ ) and split-half reliability (Spearman–Brown = .915) were strong. Convergent and discriminant validity were confirmed through the expected correlations with the Rosenberg



Self-Esteem Scale ( $r = -.577$ ), the Beck Depression Inventory-II ( $r = .397$ ), and Raven's Progressive Matrices ( $r = .028$ , ns).

**Conclusion.** The questionnaire is a reliable and valid psychodiagnostic instrument for measuring the negative impact of social media use on self-esteem and is recommended for research and applied use with Uzbek-speaking populations.

**Keywords:** Social media; self-esteem; psychometrics; reliability; construct validity; questionnaire development; social comparison; Uzbek youth

## **Introduction**

Over the past decade, social media platforms have become a pervasive feature of everyday life, particularly among adolescents and young adults, who now spend a substantial share of their waking hours engaging with image-based, algorithmically curated content. Alongside their well-documented benefits for communication, social connection, and information sharing, a growing body of research has linked intensive social media use to adverse psychological outcomes, including heightened social comparison, body image concerns, depressive symptoms, anxiety, and reduced self-esteem. The mechanisms most frequently implicated are upward social comparison, exposure to idealized and digitally edited self-presentations, and dependence on external validation expressed through likes, comments, and follower counts. Because young people are simultaneously the heaviest users of these platforms and at a developmental stage in which self-concept and identity are still consolidating, they may be especially vulnerable to these effects.

Self-esteem — the overall evaluative attitude that a person holds toward the self — is a central construct in personality and clinical psychology and a robust predictor of mental health, academic and occupational functioning, and interpersonal adjustment. Low self-esteem is consistently associated with depression, anxiety, and diminished subjective well-being. Festinger's (1954) social comparison theory provides a useful framework for understanding why social media environments may erode self-esteem: people have a fundamental drive to evaluate themselves, and in the absence of objective standards they do so by comparing themselves with others. Social media platforms intensify this



process by presenting an essentially limitless stream of selectively positive, idealized portrayals of others' appearance, achievements, and lifestyles. Such upward comparisons, repeated many times each day, are likely to produce unfavorable self-evaluations, appearance dissatisfaction, negative affect, and a fragile, externally contingent sense of self-worth.

Three interrelated mechanisms structured the construct examined in the present study. First, social comparison and low self-worth capture the tendency to feel inferior, unsuccessful, or dissatisfied with one's own life after viewing others' posts. Second, dependence on external validation reflects the extent to which a person's sense of worth becomes contingent on online feedback such as likes, comments, and follower counts — a pattern consistent with theories of contingent self-worth in which self-esteem is staked on external markers rather than internal standards. Third, body image dissatisfaction and emotional impact concern negative feelings about one's appearance relative to idealized online images, reluctance to post unedited photographs, and the deterioration of mood that can follow time spent on the platforms. Together these facets describe a coherent but multidimensional construct, suggesting that the negative impact of social media on self-esteem is unlikely to be adequately captured by a single global score.

Several established instruments measure constructs adjacent to this phenomenon, including the Rosenberg Self-Esteem Scale (Rosenberg, 1965), the Social Comparison Scale, and the Social Media Use Integration Scale. However, these measures were developed predominantly within English-speaking, Western cultural contexts, and most assess either general self-esteem or general patterns of social media engagement rather than the specific experience of self-esteem being undermined by social media. Direct translation of such instruments is methodologically problematic: linguistic equivalence does not guarantee conceptual or measurement equivalence, and culturally specific patterns of social media use, family structure, norms of self-presentation, and standards of appearance may not be adequately represented. For Uzbek youth in particular — who use social media intensively but within a distinct sociocultural and linguistic environment — no validated native-language instrument specifically targeting the negative impact of social media on self-esteem has been available. This absence constrains both research on the phenomenon in the region and the ability of practitioners to screen for and address it.

This gap motivated the development of an original Uzbek-language self-report questionnaire designed to quantify the degree of negative self-evaluation that arises from social media use. The aim of the present study was twofold: (a) to develop a culturally appropriate item pool grounded in a clearly operationalized, theoretically derived construct, and (b) to evaluate the resulting instrument against the full range of contemporary psychometric criteria, including content validity, internal consistency, factorial (construct) validity established through both exploratory and confirmatory factor analysis, convergent and discriminant validity, test–retest stability, and split-half reliability. We hypothesized that the instrument would (1) demonstrate adequate content validity as judged by expert raters, (2) exhibit a coherent three-factor structure reflecting the theorized components of the construct, (3) show high internal consistency and temporal stability, and (4) correlate negatively with general self-esteem, positively with depressive symptoms, and negligibly with nonverbal intelligence, thereby establishing convergent and discriminant validity.

## **Methods**

### **Instrument Development**

The questionnaire was developed in accordance with the psychometric methodology recommended by Anastasi and Urbina (2001), Kline (1986), and Burlachuk (2008). Development proceeded through seven sequential stages.

**Stage 1: Theoretical foundation and operationalization.** A review of the literature was used to define the target construct — “negative self-evaluation formed under the influence of social media” — and to specify three theoretically distinct components: (a) social comparison and low self-worth; (b) dependence on external validation; and (c) body image dissatisfaction and emotional impact.

**Stage 2: Initial item pool.** Between 8 and 11 items were written for each component, yielding an initial pool of 28 items.

**Stage 3: Expert (content) validity review.** Seven specialists in psychology rated the content relevance of each item, and the CVR (Lawshe, 1975) and I-CVI (Lynn, 1986) indices were computed.

**Stage 4: Comprehensibility testing.** Items were read by 15 members of the target population, and unclear wording was revised.

**Stage 5: Pilot study (N = 120).** Item performance was examined; 10 items with low discrimination (item–total  $r < .30$ ) were removed or rewritten.

**Stage 6: Main psychometric evaluation (N = 1,605).** Internal consistency, factor structure (EFA and CFA), test–retest reliability, split-half reliability, and convergent and discriminant validity were assessed.

**Stage 7: Final scale and standardization.** The 18 items with the strongest psychometric properties were retained, and percentile-based norms and interpretive cut-offs were established.

#### Participants

The main psychometric study was conducted with 1,605 respondents. Sample size was determined following the recommendations of Kline (1986) and Comrey (1988), who suggest at least 10–20 respondents per item for factor analysis ( $18 \times 20 = 360$  as a minimum). A sample of 1,605 is classified as “very good” (Comrey, 1988) and supports stable factor-analytic estimates. Stratified sampling ensured proportional representation by age (17–20, 21–25, 26–30, 31–35 years), sex (male/female), and daily duration of social media use (< 1 hr, 1–3 hr, > 3 hr). A separate test–retest sample of 60 respondents, randomly drawn from the main sample, completed the questionnaire twice at a 4-week interval. Table 1 summarizes the demographic characteristics of the main sample.

**Table 1 Demographic Characteristics of the Main Sample (N = 1,605)**

Characteristic	Category	n (%)
Sex	Male	784 (48.8%)
	Female	821 (51.2%)
Age group (years)	17–20	438 (27.3%)
	21–25	560 (34.9%)
	26–30	382 (23.8%)
	31–35	225 (14.0%)
Daily social media use	< 1 hr	233 (14.5%)
	1–3 hr	772 (48.1%)
	> 3 hr	600 (37.4%)

## Measures

**Target questionnaire.** The instrument under evaluation comprises 18 items rated on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). Total scores range from 18 to 90, with higher scores indicating a stronger negative impact of social media use on self-esteem.

**Rosenberg Self-Esteem Scale (RSES; Rosenberg, 1965).** A widely used 10-item measure of global self-esteem, included to assess convergent validity. A negative correlation with the target questionnaire was expected.

**Beck Depression Inventory-II (BDI-II).** A standard measure of depressive symptom severity, included to assess convergent validity. A positive correlation with the target questionnaire was expected.

**Raven's Progressive Matrices.** A nonverbal measure of fluid intelligence, included to assess discriminant validity. A negligible correlation with the target questionnaire was expected, as the construct is unrelated to intellectual ability.

## Procedure

Participation was anonymous and voluntary, and respondents were informed that the data would be used for research purposes only in aggregated form. The questionnaire was administered together with the validation measures. Participants were at least 17 years of age. The test–retest subsample completed the questionnaire on two occasions separated by four weeks.

## Statistical Analysis

Content validity was quantified using the CVR, I-CVI, and the average scale-level content validity index (S-CVI/Ave). Item analysis examined item means, standard deviations, corrected item–total correlations, and Cronbach's alpha if item deleted. Internal consistency was estimated with Cronbach's alpha for the total scale and each subscale. Construct validity was examined through exploratory factor analysis (principal component analysis with Varimax rotation; SPSS 26.0) and confirmatory factor analysis (AMOS 26.0). Sampling adequacy was evaluated with the Kaiser–Meyer–Olkin (KMO) index and Bartlett's test of sphericity. Model fit was assessed with  $\chi^2/df$ , CFI, TLI, RMSEA, and SRMR,

using the criteria of Hu and Bentler (1999). Convergent and discriminant validity were evaluated using Pearson correlations, and test–retest and split-half (Spearman–Brown corrected) reliability were computed.

## Results

### Content Validity

Seven experts (three Doctors of Science in psychology, three holders of the Candidate of Science / PhD degree, and one practicing psychologist with more than 10 years of experience) rated each item on a 3-point relevance scale (1 = not necessary, 2 = useful but not essential, 3 = essential). For seven raters, the adopted acceptance thresholds were  $CVR \geq .62$  (Wilson et al., 2012),  $I-CVI \geq .78$  (Lynn, 1986), and  $S-CVI/Ave \geq .90$  (Polit & Beck, 2006). All 18 retained items met these criteria, with CVR values ranging from .71 to 1.00 and I-CVI values from .86 to 1.00. The overall S-CVI/Ave was .946, indicating excellent content validity. Table 2 reports the per-item indices.

Table 2 Content Validity Indices per Item (Seven Experts)

Item	n (essential)	CVR	I-CVI	Item	n (essential)	CVR	I-CVI
1	7	1.00	1.00	10	7	1.00	1.00
2	7	1.00	1.00	11	7	1.00	1.00
3	6	0.71	0.86	12	7	1.00	1.00
4	7	1.00	1.00	13	6	0.71	0.86
5	7	1.00	1.00	14	7	1.00	1.00
6	6	0.71	0.86	15	6	0.71	0.86
7	6	0.71	0.86	16	7	1.00	1.00
8	7	1.00	1.00	17	6	0.71	0.86
9	6	0.71	0.86	18	7	1.00	1.00

*Note.* CVR = Content Validity Ratio; I-CVI = Item Content Validity Index. All items exceeded the acceptance thresholds ( $CVR \geq .62$ ;  $I-CVI \geq .78$ ).  $S-CVI/Ave = .946$ .

### Item Analysis

Item quality was evaluated through item means, standard deviations, corrected item–total correlations ( $r_{it}$ ), and Cronbach’s alpha if item deleted (Field, 2009).

All 18 items showed corrected item–total correlations between .448 and .642, well above the .30 acceptance threshold. Removing any single item did not increase the total alpha (all alpha-if-deleted values were below the overall alpha of .899), indicating that every item contributed to the scale and was retained. Table 3 presents the item statistics.

**Table 3** *Item Statistics (N = 1,605)*

Item	M	SD	r–it	$\alpha$ if deleted
1	2.85	1.12	.590	.893
2	3.00	1.15	.599	.892
3	2.50	1.12	.459	.897
4	2.85	1.19	.564	.894
5	3.15	1.14	.642	.891
6	2.76	1.13	.523	.895
7	2.40	1.11	.471	.896
8	2.97	1.16	.530	.895
9	2.18	1.07	.448	.897
10	2.69	1.13	.521	.895
11	3.04	1.13	.600	.892
12	2.76	1.12	.567	.893
13	2.58	1.13	.554	.894
14	2.88	1.13	.553	.894
15	2.62	1.12	.528	.895
16	3.08	1.15	.637	.891
17	2.46	1.10	.449	.897
18	2.76	1.15	.565	.894

*Note.* r–it = corrected item–total correlation. Acceptance criterion:  $r\text{--}it \geq .30$  (Field, 2009).

### Internal Consistency

Cronbach’s alpha for the full 18-item scale was .899, indicating good internal consistency (Nunnally, 1978; George & Mallery, 2003). Subscale reliabilities were also satisfactory to good: .864 for Factor 1 (body image dissatisfaction and emotional impact, 8 items), .848 for Factor 2 (social comparison and low self-worth, 5 items), and .783 for Factor 3 (dependence on external validation, 5 items). Table 4 summarizes these values.

**Table 4** Cronbach's Alpha for the Total Scale and Subscales

Scale	Items	$\alpha$	Level
Total scale	18	.899	Good
F1 – Body image & emotional impact	8	.864	Good
F2 – Social comparison	5	.848	Good
F3 – Dependence on external validation	5	.783	Satisfactory

### Construct Validity: Exploratory Factor Analysis

The suitability of the data for factor analysis was confirmed by the Kaiser–Meyer–Olkin index (KMO = .944, “marvellous”; Kaiser, 1974) and Bartlett’s test of sphericity,  $\chi^2(153) = 10,714.34$ ,  $p < .001$ . Applying the Kaiser criterion (eigenvalue  $> 1$ ) and Cattell’s scree test, three factors were extracted, with eigenvalues of 6.69, 1.86, and 1.39. Together they accounted for 55.18% of the total variance (F1 = 22.14%, F2 = 18.21%, F3 = 14.83%), exceeding the recommended 50% threshold (Hair et al., 2010). All primary loadings exceeded .40 (Stevens, 2002), and all cross-loadings were below .30, indicating a clear and well-differentiated factor structure. Table 5 presents the rotated factor loadings.

**Table 5** Rotated Factor Loadings (Varimax)

Item	F1	F2	F3	Assigned
1	.209	.727	.194	F2
2	.193	.772	.181	F2
3	.155	.133	.705	F3
4	.153	.746	.198	F2
5	.741	.204	.201	F1
6	.175	.732	.100	F2
7	.136	.216	.657	F3
8	.168	.243	.704	F3
9	.586	.154	.074	F1
10	.681	.172	.080	F1
11	.202	.752	.195	F2
12	.649	.208	.173	F1
13	.666	.192	.145	F1
14	.240	.211	.690	F3
15	.645	.172	.141	F1
16	.755	.196	.180	F1
17	.108	.206	.661	F3
18	.680	.196	.141	F1

*Note.* F1 = body image dissatisfaction and emotional impact; F2 = social comparison and low self-worth; F3 = dependence on external validation. Primary loadings appear in the “Assigned” column. All primary loadings  $\geq .40$ ; all cross-loadings  $< .30$ .

### Construct Validity: Confirmatory Factor Analysis

To confirm the EFA solution, a three-factor confirmatory model was estimated. All fit indices met the criteria recommended by Hu and Bentler (1999), supporting the hypothesized structure. Table 6 reports the model fit indices.

**Table 6** *Confirmatory Factor Analysis Fit Indices*

Index	Acceptance criterion	Obtained value
$\chi^2/df$	$< 3.0$ (good); $< 5.0$ (acceptable)	2.87
CFI	$\geq .90$ (acceptable); $\geq .95$ (excellent)	.953
TLI	$\geq .90$	.944
RMSEA	$\leq .08$ (acceptable); $\leq .05$ (excellent)	.058 [.052, .064]
SRMR	$\leq .08$	.047

*Note.* RMSEA reported with 90% confidence interval. The model fits the data well, providing empirical support for the three-factor structure.

### Convergent and Discriminant Validity

As predicted, total scores correlated negatively with the Rosenberg Self-Esteem Scale,  $r = -.577$ ,  $p < .001$ , and positively with the Beck Depression Inventory-II,  $r = .397$ ,  $p < .001$ , supporting convergent validity. Discriminant validity was supported by a negligible, nonsignificant correlation with Raven’s Progressive Matrices,  $r = .028$ ,  $p = .267$ , confirming that the questionnaire does not measure intellectual ability. Table 7 summarizes these correlations.

**Table 7** *Convergent and Discriminant Validity Correlations*

Measure	Type	r	p
Rosenberg Self-Esteem Scale	Convergent (expected -)	-.577	< .001
Beck Depression Inventory-II	Convergent (expected +)	.397	< .001
Raven’s Progressive Matrices	Discriminant (expected $\approx 0$ )	.028	.267

### Test–Retest and Split-Half Reliability

Temporal stability was assessed in the test–retest subsample ( $n = 60$ ) over a 4-week interval. The Pearson correlation between the two administrations was  $r = .927$ ,  $p < .001$ , far exceeding the .70 criterion (Nunnally, 1978) and indicating high stability over time. Split-half reliability was computed by dividing the items into odd and even halves; the correlation between halves was  $r = .844$ , and the Spearman–Brown corrected coefficient was .915, well above the .80 criterion. Together these results confirm that the items reliably measure a single underlying construct.

### Normative Data

Descriptive statistics for the total score were  $M = 49.51$ ,  $SD = 12.34$ , with a minimum of 18 and a maximum of 86. Skewness (0.06) and kurtosis ( $-0.39$ ) fell within  $\pm 1$ , indicating an approximately normal distribution (Kolmogorov–Smirnov test,  $p > .05$ ). Percentile-based norms divided respondents into three interpretive levels using the 33rd and 67th percentiles. Table 8 presents the normative ranges and interpretation.

**Table 8** Normative Ranges and Interpretation ( $N = 1,605$ )

Score range	Percentile	Level	Interpretation
18–43	$\leq 33\%$	Low	Minimal negative impact; psychologically resilient toward online content.
44–55	34–66%	Moderate	Moderate negative impact; monitoring of usage time and content filtering recommended.
56–90	$\geq 67\%$	High	Substantial negative impact; psychological counselling and reduced use recommended.

*Note.* Distribution of the main sample across levels: Low  $n = 558$ , Moderate  $n = 534$ , High  $n = 513$ .

### Discussion

The present study developed and comprehensively validated an original Uzbek-language self-report questionnaire measuring the negative impact of social media use on self-esteem. Across the full set of contemporary psychometric criteria, the

18-item instrument performed well, and all four study hypotheses were supported.

The instrument demonstrated excellent content validity, with an S-CVI/Ave of .946 and all item-level indices above accepted thresholds. This indicates that experts judged the items to be relevant and representative of the target construct, a necessary foundation for subsequent psychometric evaluation. The item analysis further showed that all 18 items discriminated adequately, with corrected item–total correlations clearly above .30 and no item whose removal would improve overall reliability.

Construct validity was established through convergent evidence from both exploratory and confirmatory factor analysis. The EFA produced the three theoretically anticipated factors — body image dissatisfaction and emotional impact, social comparison and low self-worth, and dependence on external validation — explaining 55.18% of the variance with a clean simple structure in which every item loaded strongly on its intended factor (all primary loadings  $\geq$  .40) and weakly on the others (all cross-loadings  $<$  .30). The CFA, estimated on the same model, corroborated this solution, with fit indices in the good-to-excellent range ( $\chi^2/df = 2.87$ ; CFI = .953; TLI = .944; RMSEA = .058 with a narrow 90% confidence interval; SRMR = .047). The convergence of EFA and CFA evidence on the same three-factor model provides strong support for the multidimensional conceptualization of the construct and suggests that the negative impact of social media on self-esteem is best understood not as a single undifferentiated dimension but as a set of related yet distinguishable facets. This structure also has practical value, because it allows clinicians and researchers to profile which aspect of the experience — appearance-related distress, comparison-driven inferiority, or validation dependence — is most prominent for a given individual, rather than relying solely on a total score.

The pattern of convergent and discriminant correlations was fully consistent with theoretical expectations. The moderate-to-strong negative correlation with the Rosenberg Self-Esteem Scale ( $r = -.577$ ) indicates that higher scores reflect lower global self-esteem, while the positive correlation with depressive symptoms ( $r = .397$ ) situates the construct within the broader network of negative affective functioning. Importantly, the negligible correlation with Raven's Progressive Matrices ( $r = .028$ ) demonstrates that the questionnaire is not confounded with intellectual ability, supporting its discriminant validity. These relationships are



consistent with social comparison theory and with prior findings linking social media use to appearance concerns and reduced well-being.

The instrument also proved highly reliable. Internal consistency was good for the total scale ( $\alpha = .899$ ) and ranged from satisfactory to good across subscales (.783–.864), the lower value for the external-validation subscale being acceptable for a relatively short five-item factor. Temporal stability was high (test–retest  $r = .927$  over four weeks), and split-half reliability was strong (Spearman–Brown = .915), indicating that scores are both stable over time and internally coherent.

Taken together, these findings can be integrated within a social comparison framework. The strong loadings on the comparison and body image factors are consistent with the view that repeated upward comparison with idealized online content drives unfavorable self-evaluation, while the external-validation factor reflects the contingent self-worth that can develop when esteem is staked on platform feedback. The fact that the three facets cohere into a single higher-order phenomenon — yet remain empirically separable — mirrors the way the literature describes social media’s influence on the self as multiply determined rather than monolithic. The negligible association with fluid intelligence further clarifies the boundaries of the construct, indicating that vulnerability to social media’s effects on self-esteem is independent of cognitive ability and is instead an affective–evaluative phenomenon.

From an applied perspective, the percentile-based norms allow practitioners to classify individuals into low, moderate, and high levels of negative impact and to tailor recommendations accordingly, ranging from preventive guidance on usage time and content filtering for those at the moderate level to referral for psychological counselling and structured work on self-esteem for those at the high level. Because the subscales are interpretable in their own right, the instrument can also support more targeted intervention — for example, cognitive work on comparison tendencies, media-literacy approaches to idealized imagery, or strategies to decouple self-worth from online feedback. Crucially, because the instrument was developed directly in Uzbek and grounded in the cultural and linguistic experience of the target population, it offers a measurement tool that translated Western scales cannot fully replicate, and it provides a basis for future epidemiological and intervention research in the region.

Several limitations should be acknowledged. First, the data are based on self-report and may be subject to social desirability and common-method biases.

Second, although the sample was large and stratified, it was drawn from a specific age range (17–35 years), and generalization to other age groups should be made with caution. Third, the test–retest subsample, while adequate, was modest in size. Future research should examine measurement invariance across sex, age, and usage groups, evaluate criterion validity against behavioral or longitudinal outcomes, and explore the instrument’s sensitivity to change in intervention contexts.

In conclusion, the developed questionnaire meets the full range of international psychometric standards for content validity, construct validity, convergent and discriminant validity, internal consistency, and temporal stability. It constitutes a reliable and valid psychodiagnostic instrument for measuring the negative impact of social media use on self-esteem and is recommended for both research and applied use with Uzbek-speaking populations.

## References

1. Anastasi, A., & Urbina, S. (2001). *Psychological testing* (7th ed.). Prentice Hall.
2. Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. Psychological Corporation.
3. Burlachuk, L. F. (2008). *Psychodiagnostics: A textbook for higher education institutions* (2nd ed.). Piter.
4. Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology, 56*(5), 754–761. <https://doi.org/10.1037/0022-006X.56.5.754>
5. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. <https://doi.org/10.1007/BF02310555>
6. Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*(2), 117–140. <https://doi.org/10.1177/001872675400700202>
7. Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). SAGE.
8. George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference* (4th ed.). Allyn & Bacon.
9. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Pearson.

10. Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
11. Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31–36. <https://doi.org/10.1007/BF02291575>
12. Kline, P. (1986). *A handbook of test construction: Introduction to psychometric design*. Methuen.
13. Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
14. Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382–385. <https://doi.org/10.1097/00006199-198611000-00017>
15. Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.
16. Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489–497. <https://doi.org/10.1002/nur.20147>
17. Raven, J. C. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology*, 41(1), 1–48. <https://doi.org/10.1006/cogp.1999.0735>
18. Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton University Press.
19. Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Lawrence Erlbaum.
20. Wilson, F. R., Pan, W., & Schumsky, D. A. (2012). Recalculation of the critical values for Lawshe's content validity ratio. *Measurement and Evaluation in Counseling and Development*, 45(3), 197–210. <https://doi.org/10.1177/0748175612440286>.