Asian Journal of Research in Education and Social Sciences
e-ISSN: 2682-8502 | Vol. 7, No. 3, 438-452, 2025
https://asianscholarsnetwork.com/asnet-journals

Assessment of Validity, Reliability, and Normality in Quantitative Study: A Survey Instrument Analysis with IBM SPSS

Nurul Hidayana Mohd Noor[1*], Amirah Mohamad Fuzi[1]

[1] Faculty of Administrative Science and Policy Studies, Universiti Teknologi MARA (UiTM), Malaysia

*Corresponding Author: hidayana@uitm.edu.my

Received: 1 February 2025 | Accepted: 17 April 2025 | Published: 30 April 2025

DOI: https://doi.org/10.55057/ajress.2025.7.3.37

Abstract: *This concept paper discusses three critical notions in a quantitative study: validity, reliability, and normality. Quantitative methods are used in the social sciences to study society and human behavior. Quantitative methods in social science include various fields, including economics, education, political science, public policy, psychology, and sociology. One of the research approaches used in quantitative study is a survey. The survey is a set of questions intended to gather insight into people's thoughts, feelings, and behaviors. An excellent survey instrument is achieved when it can fulfill the assumption of validity, reliability, and normality of the data. The validity of a survey instrument measures the extent to which it measures what it is supposed to measure. While the reliability of a survey instrument measures the extent to which the instrument is consistent in measuring a concept. Finally, the normality test helps determine that the data that has been collected is normally distributed or taken from a normal population. Good data can explain the results obtained from the study. A good measurement should have the property of isomorphism with reality and the value obtained from the measurement. Therefore, a measuring instrument is considered good when the results can accurately reflect the reality of the phenomenon to be measured.*

Keywords: Validity, Reliability, Normality, Quantitative Study, Survey

1. Introduction

Quantitative research uses quantitative information or data (Tashakkori & Teddlie, 2021). Quantitative data can be measured through a measurement process, and tools such as questionnaires and tests are required for measurement (Bloomfield & Fisher, 2019). The sample size for this study is more significant than qualitative research (Ahmad et al., 2019). Based on the perspective of the purpose, quantitative research has several points, one of which aims to develop a model. This research uses theories from literature or theory studies, but it is also essential to develop hypotheses that connect with the natural phenomena to be studied (Hodge, 2020; Jamieson et al., 2023). Another characteristic is present to answer specific problems the author raises through research (Hodge, 2020). An essential characteristic of quantitative research is that it needs to be more result-oriented. Rather process-oriented (Tashakkori & Teddlie, 2021).

Instruments are a means of research through tests and other tools to gather data as processing material. Mohajan (2020) stated that instruments are aids chosen and used by researchers in their activities to collect data. The survey is widely used in quantitative social research (Tashakkori & Teddlie, 2021). Surveys collect data or information from respondents using questionnaires distributed directly or through intermediaries such as telephone or online media (Wankhade et al., 2022). Research using the survey method can generally be described as scientific research, and data can be derived from a sample selected by the entire population (Mohajan, 2020). The survey method is a study in which the primary source of data and information is obtained by respondents as a survey sample using a questionnaire as a data collection tool (Leong & Austin, 2023). In general, the sample used as the unit of analysis is the individual. However, other units, such as dyads, groups, companies, and even cultures, can be used as units of analysis (Kent, 2020). A survey instrument cannot be directly used but must be tested for validity first. This proves that the instrument used for measuring something is valid and worthy of use (Vu, 2021). Therefore, a validity test was performed before the research was carried out.

An instrument or test can be valid when it measures what it intends to measure (Tashakkori & Teddlie, 2021). Validity comes from the word, which means how accurate and precise a measuring tool is in measuring the measurement function (Sürücü & Maslakçi, 2020). In addition, validity is a measure that shows that the measured variable is the variable that the researcher wants to study (Vu, 2021). According to Duckett (2021), validity relates to a variable measuring what should be measured. Validity in research states the degree of accuracy of the research measuring instrument to the content measured (Sürücü & Maslakçi, 2020). A questionnaire is valid if the questions can express something that will be measured by the questionnaire (Kent, 2020; Vu, 2021).

In addition to being valid, the instrument must also be reliable. The meaning of reliability is the consistency of measurement (Mohajan, 2020). According to Quintão et al. (2020), reliability refers to an understanding that instruments used in research to obtain information can be used and trusted as a data collection tool and can reveal accurate information in the field. Duckett (2021) stated that reliability is a tool to measure questionnaire indicators of variables or constructs. A questionnaire is reliable if a person's answer to a statement is consistent or stable occasionally (Tashakkori & Teddlie, 2021). There are four types of approaches to determining reliability. The first is using the reliability approach between two observers (Inter-Rater or Inter-Observer Reliability) to evaluate and see how consistent two or more observers/assessors are in evaluating a matter. If the percentage of these two observers is within ≥86 %, then this is said to have high reliability (Leong & Austin, 2023). The second is using a retesting approach (Test-Retest) to assess whether a measurement is consistent at different times. Usually, researchers need to measure a sample/object at two different times. The assumption is that reliability is high when measurements have the same value at different times. Third, a parallel form approach evaluates whether the test results are consistent. This means that the researcher needs to construct many questions and then divide them randomly into two sets. Both sets of questions were given to the same sample group. The correlation between these two sets of questions is the same and approximates the degree of reliability. Fourth is measuring the correlation between items that measure the same aspect (Leong & Austin, 2023). The researcher will give only one form of instrument to one group of individuals, which will be given only once. When consistent results are obtained from items that measure the same aspect, it has internal consistency. Methods to measure internal consistency include using coefficient alpha (Cronbach Alpha).

Finally, the researchers also need to conduct a normality test. Normality is a fundamental statistical concept, often a basic assumption in many data analysis techniques (Flick, 2022). From hypothesis testing to linear regression, the assumption that the data follows a normal distribution allows researchers to use various powerful and efficient statistical methods (Leong

Asian Journal of Research in Education and Social Sciences
e-ISSN: 2682-8502 | Vol. 7, No. 3, 438-452, 2025
https://asianscholarsnetwork.com/asnet-journals

& Austin, 2023). There are various ways to assess normality, including visual analysis such as P-P normal graphics, Normal Q-Q, Histogram, Stem Leaf, and Box Plot, as well as statistical analysis such as Kolmogorov Smirnov, Shapiro Wilk, Shapiro Francia, Andersen Darling, Ryan Joiner, Skewness Kurtosis Test, and Jarque Bera test (Tashakkori & Teddlie, 2021). Surveys are widely used research methods to collect data from respondents to understand better a topic or phenomenon (Wankhade et al., 2022). Using surveys, researchers can access extensive and in-depth information from respondents representing a specific population (Leong & Austin, 2023; Tashakkori & Teddlie, 2021). In conclusion, good data quality requires strict data management, accurate requirements collection, comprehensive testing, and careful data pipeline design. Therefore, the paper aims to present several ways and approaches to improving the instruments used in the survey study. The researchers also provided a sample of the tests used and how to present the test's outcome. This study is essential since there is a lack of guidance for social science researchers to refer to. Moreover, past studies have focused heavily on validity and reliability, while the normality assumption has been discussed separately.

2. Literature Review

Validity

Validity is one of the most important criteria and should be considered when evaluating the test (Allan & Skinner, 2020). According to Chapelle and Voss (2021), validity refers to the extent to which a tool measures what it should measure. Nayak and Singh (2021) explained that validity refers to the extent to which the measurement is helpful in its purpose. Chapelle and Lee (2021) asserted that a measurement has high validity if the degree of ability to measure what is supposed to be measured is high. Several factors may influence validity tests. Among the factors that affect validity is 1) instructions are not clear, 2) inappropriate sentence structure and language, 3) the level of difficulty of the question is not suitable, 4) bad question construction, 5) ambiguity, 6) questions are not appropriate for the skill level being measured, 7) lack of time, 8) the test is too short, 9) incorrect order of questions, and 10) answer patterns that can be identified by students (Chapelle & Lee, 2021; Chapelle & Voss, 2021; Nayak & Singh, 2021). There are several types of validity (see Figure 1).

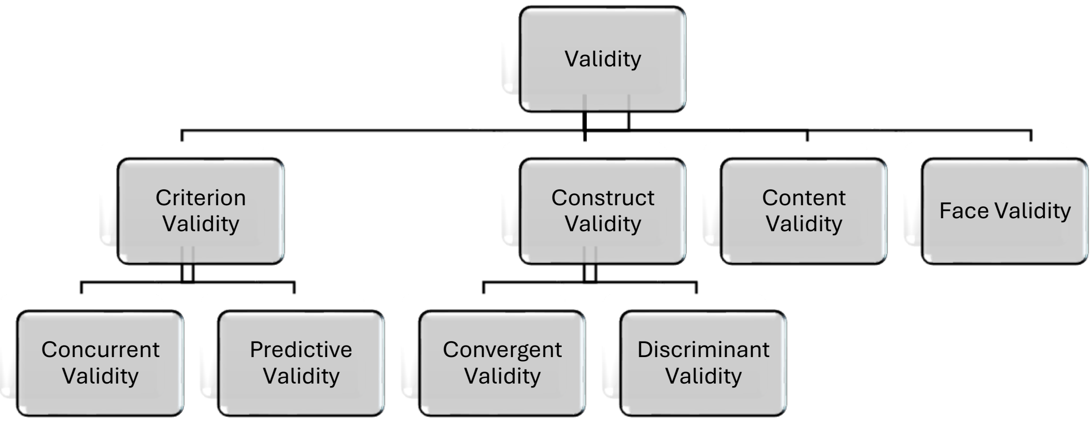


Figure 1: Types of Validity

First is content validity. Content validity is a type of validity that assesses whether a test represents all aspects of the construct (Madadizadeh & Bahariniya, 2023). To produce valid results, the content of the test, survey, or measurement method must cover all relevant parts of the research subject that is to be measured. For example, a math teacher develops an end-of-semester algebra test for her class. The test should cover every form of algebra taught in class.

Asian Journal of Research in Education and Social Sciences
e-ISSN: 2682-8502 | Vol. 7, No. 3, 438-452, 2025
https://asianscholarsnetwork.com/asnet-journals

If some types of algebra are omitted, then the results may not accurately indicate students' understanding of a particular research subject. Similarly, if she included algebra-related questions, the results would no longer be a valid measure of algebraic knowledge. Content validity ensures that the measurement includes adequate and representative items that reveal the concept being studied (Madadizadeh & Bahariniya, 2023).

Examples of Content Validity Test

The selection of three experts is made to verify the content of the research instrument by looking at their experience and expertise in organizational studies. After the expert review, agreement scores for face validity and content validity of the survey questions among the experts should be determined using the Content Validity Index (CVI). The average rating for each construct and the degree of appropriateness will be determined through the CVI count from each expert. A good CVI value is ≥ 0.80 (Davis, 1992; Polit et al., 2007). The formula for CVI count is as follows:

$$\text{Content Validity Index (CVI)} = \frac{\text{Total score of each expert}}{\text{Actual total score}}$$

The instrument's content validity is determined by the CVI value (Nayak & Singh, 2021). Table 1 shows the expert assessment related to the content of the readiness instrument in this study. Three experts reviewed the content validity of this instrument. A total of 18 items have been checked and rated for each item.

Table 1: Item Content Validity

Item	Expert 1	Expert 2	Expert 3
AC1	1	1	1
AC2	1	1	1
AC3	1	1	1
AC4	0	1	1
AC5	1	1	1
AC6	1	1	1
CC1	1	1	1
CC2	1	1	1
CC3	1	1	1
CC4	1	1	1
CC5	1	1	1
CC6	1	1	0
NC1	1	1	1
NC2	1	1	1
NC3	1	1	1
NC4	1	1	1
NC5	1	1	1
NC6	1	1	1
Score	17	18	17

The results in Table 1 show that Expert One has agreed on 17 items out of 18 items, Expert Two agreed on as many as 18 items, while the relevant items for Expert Three are as many as 17. The view of Polit and Beck (2006) was used to calculate the CVI. Value CVI for expert one - 17 items / 18 items, expert two - 18 items / 18 items, and expert three - 17 items / 18 items. Table 2 shows the CVI value obtained for each expert according to construct and value

Asian Journal of Research in Education and Social Sciences
e-ISSN: 2682-8502 | Vol. 7, No. 3, 438-452, 2025
https://asianscholarsnetwork.com/asnet-journals

overall CVI. Overall, the average content validity index according to the CVI value obtained for this study is 0.962. The results show that this research instrument has content validity.

Table 2: CVI Value By Construct

Construct	Expert 1	Expert 2	Expert 3	Average CVI
Affective Commitment	0.833	1	1	0.944
Continuance Commitment	1	1	0.833	0.944
Normative Commitment	1	1	1	1
Overall CVI value				**0.962**

Next, face validity is the most basic and minimal content validity (Sekaran & Bougie, 2010). The previous explanation is similar, Groth-Marnat (2009) explained that content and face validity have differences and are not synonymous. Content validity concerns the judgment made by experts, while face/appearance validity concerns the judgment of test users. In line with that, Gregory et al. (2020) explained that face validity is the level of acceptance by people in general towards the measurement function of the test and is not related to validity statistics such as coefficients or indexes. Face validity is like content validity, but face validity is a more informal and subjective assessment (Sekaran & Bougie, 2010). Because face validity is a subjective measure, it is often considered the weakest form of validity. However, this can be useful in the early stages of research method development.

Criterion validity emphasizes whether the results are consistent with different tests for the same thing; in other words, criterion validity evaluates how closely our test results match the results of different tests (Flake et al., 2022). The term criterion itself refers to an external measure of the same thing. These are usually well-established or widely used tests that are already considered valid. To evaluate the validity, we must calculate the correlation between the measurement results and the measurement results of the criteria. If there is a high correlation, this gives a good indication that our test is measuring what it wants to measure. The difference between the types of validity related to different criteria is in the criteria used as a standard for evaluation. Specifically, the types of validity related to the criteria can be divided into concurrent and predictive validity (Nayak & Singh, 2021). In concurrent validity, we assess the operationalization's ability to differentiate between groups that theoretically should be distinguishable.

In predictive validity, we evaluate the operationalization's ability to predict something that theoretically should be predictable (Sekaran & Bougie, 2010). For example, a measure of mathematical ability should be able to predict how well a person will perform in a technical-based profession. We can give the measurements we have made to experienced technicians and see if there is a high correlation between the scores on the measurements and their salaries as technicians. A high correlation will provide evidence for the predictive validity of this and show that our measure can correctly predict something that we think theoretically should be predictable (Flake et al., 2022). Finally, construct validity refers to the extent to which a test can measure hypothetical constructs about human nature and behavior (Sekaran & Bougie, 2010). It can be categorized into two, which are convergent validity and discriminant validity. In convergent validity, we examine the extent to which an operationalization is like (converses on) other operationalizations that theoretically should be similar (Flake et al., 2022). For example, to show the convergent validity of an arithmetic skill test, we might correlate the score on our test with the score on another test intended to measure essential mathematical ability, where a high correlation would be evidence of convergent validity. In discriminant validity, we examine the extent to which an operationalization is not the same as (deviates from) other operationalizations that theoretically should not be similar (Nayak & Singh, 2021). For example, to show the discriminant validity of an arithmetic skills test, we might correlate our test scores with verbal ability test scores, where a low correlation would be evidence of discriminant validity.

Validation of Research Instrument with Exploratory Factor Analysis (EFA)

Construct validity is a test to see if the items in the research instrument are suitable to measure the existing theoretical construct. The validation procedure from this factor analysis is also called factorial validity (Shrestha, 2021). Factor analysis is done to see if the present items represent aspects or dimensions that should be measured (Shrestha, 2021). In addition, factor analysis is also performed to show whether the aspects or dimensions are interrelated or not (independent). If, at the initial development stage of the measurement tool, the researcher usually does not determine the number of factors, then in the validation procedure, the researcher has determined the number of factors to be extracted by the existing theory (Goretzko et al., 2021). In addition, the minimal loading factor obtained from the item against the aspect or dimension measured is also expected to be high. Stevens (1992) recommends that only items with a loading factor above 0.4 are worthy of retention. The following is an example of exploratory factor analysis to see if the Organizational Commitment scale has satisfactory factorial validity. Theoretically, the most accepted tool to measure organizational commitment is that of Allen and Meyer (1990).

Table 3: Organizational Commitment Scale

Organizational Commitment	Items
Affective Commitment	AC1. I would happily spend the rest of my career in this organization.
	AC2. This organization's problems are my own.
	AC3. I do not feel like a "part of my family" at this organization.
	AC4. I do not feel “emotionally attached” to this organization.
	AC5. This organization has a great deal of personal meaning for me.
	AC6. I do not feel a strong sense of belonging to this organization.
Continuance Commitment	CC1. It would be tough for me to leave my job at this organization right now, even if I wanted to.
	CC2. More of my life would be disrupted if I stayed in my organization.
	CC3. Currently, staying with my job at this organization is a matter of necessity as much as desire.
	CC4. I have too few options to consider leaving this organization.
	CC5. One of the few negative consequences of leaving my job at this organization would be the scarcity of available alternatives elsewhere.
	CC6. One of the major reasons I continue to work for this organization is that leaving would require considerable personal sacrifice.
Normative Commitment	NC1. I do not feel any obligation to remain with my organization.
	NC2. Even if it were to my advantage, leaving would be wrong.
	NC3. I would feel guilty if I left this organization now.
	NC4. This organization deserves my loyalty.
	NC5. I would not leave my organization right now because of my sense of obligation to it.
	N6. I owe a great deal to this organization.

To perform factor analysis, follow these steps:

1) Analyze – Dimension Reduction – Factor
2) Enter all items into variables
3) Click descriptive and tick KMO and Bartlet's test of sphericity - click Continue.
4) Click Extraction, tick Scree Plot, and then click Continue. Then, because we set the number of factors based on theory, select a factor to extract and fill in 3 (according to theory)

Asian Journal of Research in Education and Social Sciences
e-ISSN: 2682-8502 | Vol. 7, No. 3, 438-452, 2025
https://asianscholarsnetwork.com/asnet-journals

5) Click Rotation and select Varimax.
6) Click options, check sorted size, and suppress small coefficient. Determine the absolute value below 0.3.
7) Click OK

The first output is the KMO table and Bartlett's test. The basic assumption of exploratory factor analysis is that each item is correlated. A sufficient sample is needed to see whether there is a correlation. Therefore, the Kaiser-Meyer-Olkin (KMO) value shows the sample's adequacy. Generally, a KMO value above 0.5 indicates a sufficient sample. Then Bartlett's sphericity test shows a correlation between variables; if significant (sig <0.05), then factor analysis can be done. The output of the analysis results shows a KMO value of 0.887, and Bartlett's test of sphericity is significant. Therefore, the factor analysis can be continued. The communalities table shows the effective contribution of each item to the factors formed. For example, in item 1, the item gives an effective contribution of 61.8% to the factor formed; it can be said that this item is good enough to explain the variance in the factor.

Table 4: KMO and Bartlett's Test

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.887
Bartlett's Test of Sphericity	Approx. Chi-Square	4280.586
	df	300
	Sig.	0.000

Communalities

	Initial	**Extraction**
NC1	1.000	0.618
NC2	1.000	0.610
NC3	1.000	0.674
NC4	1.000	0.602

The following output is a table of total variance explained. Total variance explained is the percentage of variance of the measurement construct that can be explained by several factors that are formed. Because we have decided to extract into three factors at the beginning, the variance is only explained up to the 3rd factor. From the initial eigenvalues column in the cumulative variance, reducing 18 items to one factor can explain 26.06% of the variance, reducing it to two factors can explain 39.21% of the variance, and reducing it to three factors can explain 47.25%.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6,517	26,068	26,068	6,517	26,068	26,068	4,396	17,582	17,582
2	3,286	13,142	39,210	3,286	13,142	39,210	4,238	16,953	34,535
3	2,010	8,040	47,250	2,010	8,040	47,250	3,179	12,715	47,250
4	1,312	5,248	52,498						
5	1,044	4,178	56,676						

Figure 2: Total Variance Explained

We can reduce the many factors from the pattern on the scree plot. As seen from the picture, after the third point, the line begins to experience a change in slope, and the variation explained is getting less and less. Thus, we can reduce the 18 items to only three factors, and this is to the existing theory.

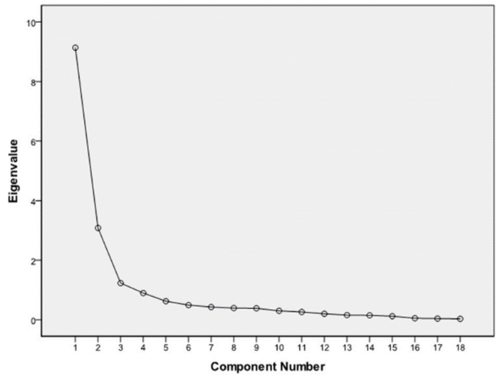


Figure 3: Scree Plot

The following output seen is the rotated component matrix. Table 4 shows the loading factor on each factor. From Table 5, the items are generally well distributed and grouped according to their factors. Stevens (1992) suggests using items that have a loading factor above 0.4 while the item has an item below 0.4. In addition, if the item has a high loading factor on both factors. The item measures two measurement dimensions. Justification is required to eliminate this item. In deciding whether an item will be eliminated, the consideration used is the balance of the item composition in aspects or dimensions, professional justification based on existing theory, and the reliability produced whether the item will be eliminated.

Table 5: Rotated Component Matrix

Item	Factor		
	1	2	3
AC1		0.852	
AC2		0.868	
AC3		0.813	
AC4		0.808	
AC5		0.799	
AC6		0.756	
CC1	0.716		
CC2	0.780		
CC3	0.799		
CC4	0.726		
CC5	0.709		
CC6	0.711		
NC1			0.773
NC2			0.813
NC3			0.823
NC4			0.794
NC5			0.765
NC6			0.811

Reliability

Reliability is often referred to as internal stability and consistency. Cronbach's Alpha value is often referred to as today's measurement of a construct's internal consistency. Scholars have revealed that the sample sizes of 20 and 30 using the Cronbach alpha statistics needed reliability (Nunnally & Bernstein, 1994). The instrument's reliability became more robust when the sample size was at least 100. Following are examples of scholarly views on Cronbach's Alpha value:

1) According to Babbie (1992), Cronbach Alpha values are classified based on the classification in which the reliability index of 0.90-1.00 is very high, 0.70-0.89 is high, 0.30-0.69 is moderate, and 0.00 to 0.30 is low.
2) For Sekaran (1992), a reliability value of less than 0.60 is considered low and unacceptable; an alpha value between 0.60 and anand0.80 is acceptable, while an alpha value that exceeds 0.80 is considered good.
3) According to Kline (2000), a reliability of 0.7 is a minimum for a good test.
4) A Cronbach Alpha value exceeding 0.60 is often used as an index of an instrument's reliability (Pallant, 2001).
5) Nunnally and Bernstein (1994) suggested that a Cronbach's alpha value of 0.70 or higher demonstrates satisfactory reliability.
6) Hair et al. (2016) stated that the value of the reliability analysis could be interpreted based on the strength as follows: < 0.6 = Poor, 0.6 to < 0.7= Moderate, 0.7 to < 0.8 = Good, 0.8 to < 0.9 = Very Good, and >0.9 = Excellent.

There are four ways to assess reliability. First is internal consistency, which examines the consistency of different items within the same test. Second is inter-rater (Fink & Litwin, 1995). In this method, multiple independent judges score the test on its reliability. Third is parallel or alternate forms. This approach uses different forms of the same test and compares the results. Instrument reliability is calculated by correlating the data of one instrument with the data of an equivalent instrument. The instrument can be declared reliable if the correlation is positive and significant. Fourth is test-retest (Fink & Litwin, 1995). This measures the reliability of results by administering the same test at different points in time.

Reliability of Research Instrument Using SPSS

To perform reliability analysis in SPSS, follow these steps:

1) Analyse – Scale – Reliability Analysis
2) In the left box, transfer all variables into Box Items. In the box Scale Label, write the name of your scale (e.g., Affective Commitment).
3) Click Statistics - Choose the options. Tick item, scale, and scale if the item is deleted.
4) Click continue and then OK.

A Cronbach Alpha value exceeding 0.60 is often used as an index of an instrument's reliability (Pallant, 2001). Therefore, the results indicate that the scale has good reliability and internal consistency (Cronbach's alpha coefficient = 0.757).

Normality

A normality test is performed to evaluate the distribution of data on a group of data or variables and whether the distribution of the data is normally distributed (Nor, 2015). The statistical tests that can be used include the Chi-Square Test, Kolmogorov Smirnov, Lilliefors, Shapiro Wilk, and Jarque Bera. The normality test also allows researchers to understand how much their data approaches a normal distribution so they can choose the appropriate analysis method. In general, the normality test can be done through three techniques: visual, skewness-kurtosis, and statistical tests (Kolmogorov-Smirnov and Shapiro-Wilk) (Ghasemi & Zahediasl, 2012).

Visual Method

The visual method is to look at the histogram directly. Data can be normal if the shape resembles a symmetrical inverted bell (Ghasemi & Zahediasl, 2012). If a histogram for a

Asian Journal of Research in Education and Social Sciences
e-ISSN: 2682-8502 | Vol. 7, No. 3, 438-452, 2025
https://asianscholarsnetwork.com/asnet-journals

dataset is roughly bell-shaped, then it is likely that the data is normally distributed. An example of a histogram can be seen in Figure 4.

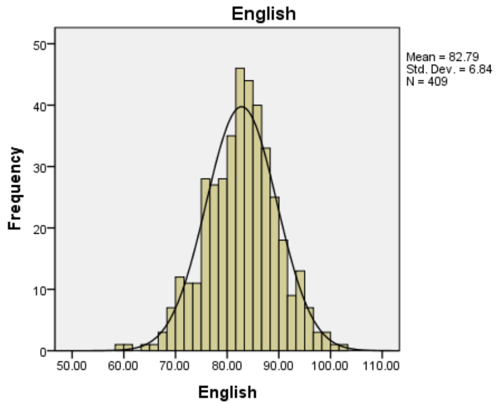
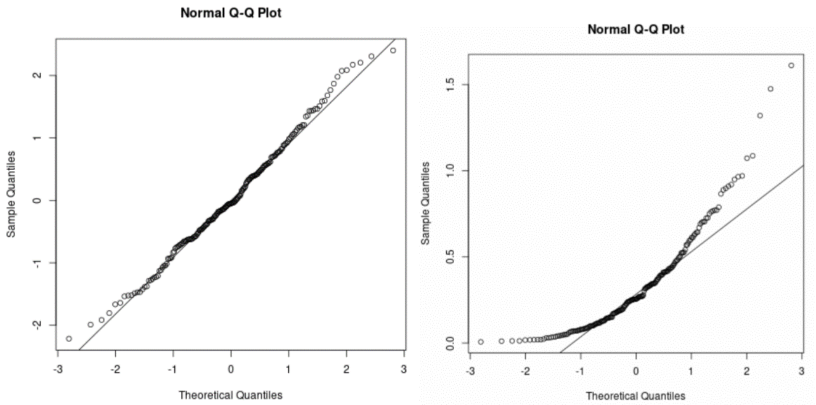


Figure 4: Histogram of Normal Distribution

In addition to displaying a histogram, the researchers can view P-P and Q-Q plots. Figure 5(a) shows that the points lie primarily along the straight diagonal line with some minor deviations along each tail. Thus, the data are normally assumed. Next, Figures 5(b) and 6 show that the points deviate significantly from the straight diagonal line. This indicates that the data set is not normally distributed.



(a) Normal (b) Nonnormal

Figure 5: Q-Q Plot

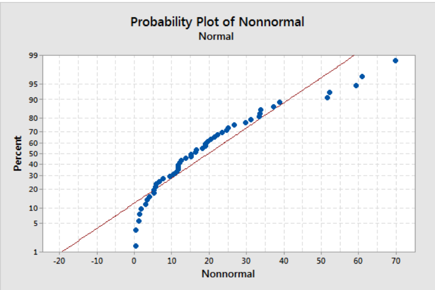


Figure 6: P-P Plot (nonnormal)

Skewness and Kurtosis

The assumption of normality of the data affects the model being fitted. The Skewness value should fall within the range of -3 to +3. Moreover, for kurtosis, the range of -10 to +10 needs to be assumed (Kline, 2005). The term 'skewness' means the absence of symmetry from the average dataset. In statistics, kurtosis is defined as the relative sharpness parameter of the peak of the probability distribution curve (Demir, 2022). This ensures that the observation method is clustered around the distribution center. This is used to show the flatness or peak of the frequency distribution curve and measure the tail or outlier of the distribution. Positive kurtosis indicates that the distribution is more peaked than the normal distribution, while negative kurtosis indicates that the distribution is less peaked than the normal distribution (Demir, 2022).

To perform skewness and kurtosis analysis in SPSS, follow these steps:

1. Analyse – Descriptive Statistics – Descriptives
2. In the new window that appears, drag the compute variable to the Variables panel
3. Then click the Options button. Check the boxes next to Kurtosis and Skewness in the new window that appears.
4. Then click Continue. Then click OK.

As shown in Table 6, the variables used are acceptable and reliable and can be used in the following study. For Skewness and Kurtosis, the variables used are normally distributed (Kline, 2005).

Table 6: Normality Results

Variable	Mean	SD	Skewness	Kurtosis
Affective Commitment	4.209	0.714	-1.179	2.888
Continuance Commitment	4.069	0.789	-1.004	1.721
Normative Commitment	4.015	0.811	-0.729	0.649

Shapiro-Wilk and Kolmogorov-Smirnov

Shapiro-Wilk and Kolmogorov-Smirnov tests are generally used for univariate data. The univariate test will test the normality of the data for each variable in the data and produce normality test results for as many variables as are tested (Khatun, 2021). The Shapiro–Wilk test is a more appropriate method for small sample sizes (<50 samples), although it can also handle larger sample sizes, while the Kolmogorov–Smirnov test is used for n ≥50.

To perform Shapiro-Wilk and Kolmogorov-Smirnov tests in SPSS, follow these steps:

1) Click Analyse > Descriptive Statistics > Explore
2) Move the variable of interest from the left box into the Dependent List box on the right.
3) Click on the SPSS Statistics Button.
4) Leave the above options unchanged and click on the SPSS Continue Button.
5) Click the SPSS Plots Button and tick the Normality plots with tests option.
6) Click on the Continue button.
7) Click on the OK button.

Asian Journal of Research in Education and Social Sciences
e-ISSN: 2682-8502 | Vol. 7, No. 3, 438-452, 2025
https://asianscholarsnetwork.com/asnet-journals

Based on the results obtained in Figure 7, the data are normally distributed if the significance value exceeds the alpha value (0.05).

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Frisbee Throwing Distance (Metres)	.091	50	.200*	.963	50	.118

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure 7: Shapiro-Wilk and Kolmogorov-Smirnov Result

4. Discussion

The most important thing to consider before designing the questionnaire is to re-examine the study's objectives. The meaning here is that the questions in the questionnaire should consider the study's objectives. Some of the steps the researcher needs to think about before developing a questionnaire include identifying the information (data) needed (to answer the research question/hypothesis), identify who the respondents of study are, choose a data collection method that suits the study respondents, think about the content (constructs and items) in the questionnaire (to be in line with the objective/research question/hypothesis), develop survey items, arrange the questions in the appropriate order, and do a pilot test of the questionnaire. Based on the feedback from the pilot test, the researchers need to make improvements/modifications to the items. To reduce the bias that arises during the actual study, Churchill (1995) suggested seven steps in doing validity and reliability, namely: 1) translation with the back-to-back technique, 2) reviews by colleagues and supervisors, 3) confirmation and comments from experts, 4) Interview, 5) assessment of item clarity, 6) assessment of internal reliability consistency, and 7) assessment of construct validity through exploratory factor analysis (EFA).

Furthermore, there are a few strategies for improving research instruments, such as questionnaires, to enhance data quality and reliability in studies. Firstly, pilot testing is conducted to identify and rectify any ambiguities, confusing questions, or issues with response options (Norland-Tilburg, 1990). Secondly, content validity ensures that the questionnaire covers all relevant aspects of the study topic by reviewing it against the study objectives and consulting subject matter experts. Thirdly, clarity and simplicity use clear and straightforward language to minimize respondent confusion and improve comprehension. Fourthly is a reliability check, assessing the questionnaire's reliability to ensure consistent results over time and across different settings. Fifth, feedback incorporation incorporates feedback from pilot testing and experts to refine the questionnaire further, enhancing its effectiveness.

5. Conclusion

An instrument is a tool used to collect data in research. The data collected using specific instruments will be described and attached or used to test the hypothesis proposed in the research. Validity is the extent to which an instrument performs its function or measures what it is supposed to measure. This means the extent to which an instrument is accurate and precise in performing its function. Reliability shows the extent to which the instrument can be trusted. The more it matches the actual score, the higher the reliability. Instrument validity and reliability are two fundamental concepts in research, and they are interrelated. With instrument validity and reliability, the research can produce meaningful findings, and the conclusions drawn from the research will be questioned. In addition, using valid and reliable instruments can lead to correct decision-making in various contexts, such as education, health, business, and other fields.

Acknowledgment
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

Ahmad, S., Wasim, S., Irfan, S., Gogoi, S., Srivastava, A., & Farheen, Z. (2019). Qualitative v/s. Quantitative research-a summarized review. *Population, 1*(2), 2828-2832.
Allan, G., & Skinner, C. (Eds.). (2020). *Handbook for research students in the social sciences*. Taylor & Francis.
Allen, N. J., & Meyer, J. P. (1990). The measurement and antecedents of affective, continuance, and normative commitment to the organization. *Journal of Occupational Psychology, 63,* 1–18.
Babbie, E. (1992). *The practice of social research*. California: Wardsworth Publishing Company
Bloomfield, J., & Fisher, M. J. (2019). Quantitative research design. *Journal of the Australasian Rehabilitation Nurses Association, 22*(2), 27–30.
Chapelle, C. A., & Lee, H. W. (2021). Conceptions of validity. In *The Routledge handbook of language testing* (pp. 17-31). Routledge.
Chapelle, C. A., & Voss, E. (Eds.). (2021). *Validity argument in language testing: Case studies of validation research*. Cambridge University Press.
Churchill, G. A., Jr. (1995). *Marketing research: Methodological foundations*. Chicago: Dryden.
Davis, L. L. (1992). Instrument review: Getting the most from your panel of experts. *Applied Nursing Research, 5*, 194–197.
Demir, S. (2022). Comparison of normality tests in terms of sample sizes under different skewness and Kurtosis coefficients. *International Journal of Assessment Tools in Education, 9*(2), 397–409.
Duckett, L. J. (2021). Quantitative research excellence: Study design and reliable and valid measurement of variables. *Journal of Human Lactation, 37*(3), 456–463.
Fink, A., & Litwin, M. S. (1995). *How to measure survey reliability and validity* (Vol. 7). Sage.
Flake, J. K., Davidson, I. J., Wong, O., & Pek, J. (2022). Construct validity and the validity of replication studies: A systematic review. *American Psychologist, 77*(4), 576.
Flick, U. (2022). *The SAGE handbook of qualitative research design*. Freie Universität Berlin, Germany.
Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism, 10*(2), 486.
Goretzko, D., Pham, T. T. H., & Bühner, M. (2021). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology, 40*, 3510-3521.
Gregory, A. J., Anderson, J. F., & Gable, S. L. (2020). You don't know how it feels: Accuracy in emotion perception predicts responsiveness of support. *Emotion, 20*(3), 343.
Groth-Marnat, G. (2009). *Handbook of psychological assessment* (5th Edition). John Wiley & Sons. Hughes, S. A., Deville, C., Chalhoub, M., & Romboletti, R. (1992). The Rorschach human anatomy response: Predicting sexual offending behavior in Juveniles. *The Journal of Psychiatry & Law, 20*(3), 313–333.

Hair Jr, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2016). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage publications.

Hodge, S. R. (2020). Quantitative research. In *Routledge Handbook of Adapted Physical Education* (pp. 147–162). Routledge.

Jamieson, M. K., Govaart, G. H., & Pownall, M. (2023). Reflexivity in quantitative research: A rationale and beginner's guide. *Social and Personality Psychology Compass*, *17*(4), e12735.

Kent, R. (2020). *Data construction and data analysis for survey research*. Bloomsbury Publishing.

Khatun, N. (2021). Applications of normality test in statistical analysis. *Open Journal of Statistics*, *11*(01), 113.

Kline, R. B. (2000). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press

Leong, F. T., & Austin, J. T. (Eds.). (2023). *The psychology research handbook: A guide for graduate students and research assistants*. SAGE publications.

Madadizadeh, F., & Bahariniya, S. (2023). Tutorial on how to calculating content validity of scales in medical research. *Perioperative Care and Operating Room Management*, 100315.

Mohajan, H. K. (2020). Quantitative research: A successful investigation in natural and social sciences. *Journal of Economic Development, Environment, and People*, *9*(4), 50-79.

Nayak, J. K., & Singh, P. (2021). *Fundamentals of research methodology problems and prospects.* SSDN Publishers & Distributors.

Nor, A. W. (2015). The graphical assessment of multivariate normality using SPSS. *Education in Medicine Journal*, *7*(2).

Norland-Tilburg, E. V. (1990). Controlling error in evaluation instruments. *Journal of Extension*, *28*(2), 23-41.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed). McGraw-Hill: New York.

Pallant, J. (2001). *SPSS survival manual -a step by step guide to data analysis using SPSS for Windows* (version 10). Buckingham Open University Press.

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 and Health*, *29*, 489–497.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Focus on research methods is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, *30*, 459–467.

Quintão, C., Andrade, P., & Almeida, F. (2020). How can we improve the validity and reliability of a case study approach? *Journal of Interdisciplinary Studies in Education*, *9*(2), 264-275.

Sekaran, U. (2003). *Research methods for business: A skill building approach* (2nd Edition). New York: John Wiley & Sons, Inc.

Sekaran, U., & Bougie, R. (2010). *Research methods for business: A skill-building approach* (5th Ed.). Haddington: John Wiley & Sons.

Shrestha, N. (2021). Factor analysis as a tool for survey analysis. *American Journal of Applied Mathematics and Statistics*, *9*(1), 4–11.

Stevens, J. P. (1992). *Applied multivariate statistics for the social sciences*. Hillsdale, NJ: L. Erlbaum Associates.

Sürücü, L., & Maslakçi, A. (2020). Validity and reliability in quantitative research. *Business & Management Studies: An International Journal, 8*(3), 2694-2726.

Tashakkori, A., & Teddlie, C. (2021). *Sage handbook of mixed methods in social & behavioral research*. SAGE publications.

Vu, T. T. N. (2021). Understanding validity and reliability from qualitative and quantitative research traditions. *VNU Journal of Foreign Studies, 37*(3).

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review, 55*(7), 5731-5780.