The Development Of Neutrosophic Form Of The Satisfaction With Life Scale And Proposal For A Confirmatory Analysis Based On Neutrosphic Logic

Volkan Duran
Iğdır Üniversites: Igdir Universitesi

Selçuk Topal (s.topal@beu.edu.tr)
Bitlis Eren University https://orcid.org/0000-0002-5829-5257

Research Article

Keywords: Convolutional Neural Netwok, Neutrosophic Logic, Scale Development, Neutrosophic Social Science, Validity

Posted Date: March 1st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-188182/v1

The Development Of Neutrosophic Form Of The Satisfaction With Life Scale And Proposal For A Confirmatory Analysis Based On Neutrosphic Logic

[1]Volkan Duran and [2] Selçuk Topal

[1]Iğdır University, Faculty of Science and Arts, Iğdır, Turkey

volkan.duran@igdir.edu.tr

[2] Bitlis Eren University, Faculty of Science and Arts, Bitlis, Turkey

s.topal@beu.edu.tr

Abstract

The main concept of neutrosophy, introduced by Smarandache (1998) is that any idea has not only a certain degree of truth but also a degree of falsity and indeterminacy in its own right. Although there are many applications of neutrosophy in different disciplines, the incorporation of its logic in education and psychology is rather scarce compared to other fields. In this study, the Satisfaction with Life Scale adapted into the Turkish by Dağlı and Baysal (2016) which was developed by Diener, Emmons, Larsen ve Griffin (1985) was converted into the neutrosophic form and the results were compared in terms of confirmatory analysis by convolutional neural networks. To sum up, two different formulas are proposed at the end of the study in order to determine the validity of any scale in terms of neutrosophy.

Keywords: Convolutional Neural Netwok, Neutrosophic Logic, Scale Development, Neutrosophic Social Science, Validity

1.Introduction

Scale development is an important part of computational social science researches, especially for quantitative researches. Therefore, these researches mostly rely on psychometric researches. Usually, psychometricians assess human differences by administering test batteries that have been found to have accurate measuring properties. Effects from these tests are then evaluated from factor analysis and multidimensional scaling to classify latent variables or factors responsible for similar trends of correlations. Specific differences for aimed cognitive skill are generally represented in terms of factors in those studies [1]. The main objective of those who support the psychometric strategy is to allow for the assessment to be made objective. From this standpoint, assessment should be based on objective determinations. For this reason, the psychometric approach emphasizes scales based on statistical methods such as factor analysis, item analysis, and test analysis, and tests its validity and reliability with scientific methods [2].

The numerical properties obtained depending on the group to which a test is applied are generally called test statistics. Some of the test statistics can be calculated based on item statistics. In general, the test statistics like the average of the test, the average difficulty of the test, the variance of the test, and other

test statistics are highly useful [3]. Researchers want to show whether there is harmony in an instrument's responses. Factor analysis is one of the multivariate approaches that social scientists use to validate psychological aspects. When several independent variables are grouped in a single study, statistical analysis can become rather challenging. It is often advantageous to group together those variables that are correlated with one another. Factor analysis is a technique that allows researchers to see whether many variables can be portrayed as a few factors [4]. Factor analysis seeks to identify some new specific factors by putting together a small number of factors that aren't connected (a p-dimensional space) [5]. It is recommended that the scale from the explanatory factor analysis process should be tested through confirmatory factor analysis [6]. Confirmatory factor analysis could be considered as a way to verify the validity of factor structures. Using this method, it is attempted to prove that the observed variables are connected with the hidden variables and hidden variables are connected. To investigate these relationships, measurement models were built [7].

There are three types of factors for developing a more grounded scale: (i) reliability; (ii) validity; and (iii) sensitivity. Reliability refers to the extent to which a measurement of a phenomenon produces consistent results [8]. Therefore, reliability means consistency or stability. Consistency of any measurement scale is important for objective scientific researches and this concept is related to 'agreement', 'reproducibility', and 'repeatability' of any measurement. The agreement is the closeness of two measurements made on the same subject as opposed to one another. Reliability includes repeatability. Repeatability means measuring accurately the same variable again and again for the same circumstances [9]. A test or measure is said to be reliable if there are always identical results using the same testing procedure [10]. This means that regardless of how many times the measurement has been taken or by whom it has been performed, you will always obtain the same value. This means two things: first of all, you should get the same result each time you use the measure, and secondly, you should use the measure as many times as possible This can be an issue in data collection when several people are involved [11]. Reproducibility referred to variations in test results while tests were performed on subjects on different occasions. The changed circumstances may be due to the use of various methods of measurement or instruments, measurement by several observers or raters, or measurements during a period in which the variable's 'error-free' level may undergo a non-negligible change [9].

Reliability is, therefore, the level of error-free. As the amount of error decreases as a result of measurement, reliability increases, and as the number of errors increases, reliability decreases. Reliability levels of measurement tools are determined by reliability analysis. Reliability is best expressed with the reliability coefficient (r) ranging from 0.00 to +1.00. The closer the reliability coefficient of the measurement tool is to 1, the higher the reliability, the closer to 0, the lower the reliability [12].

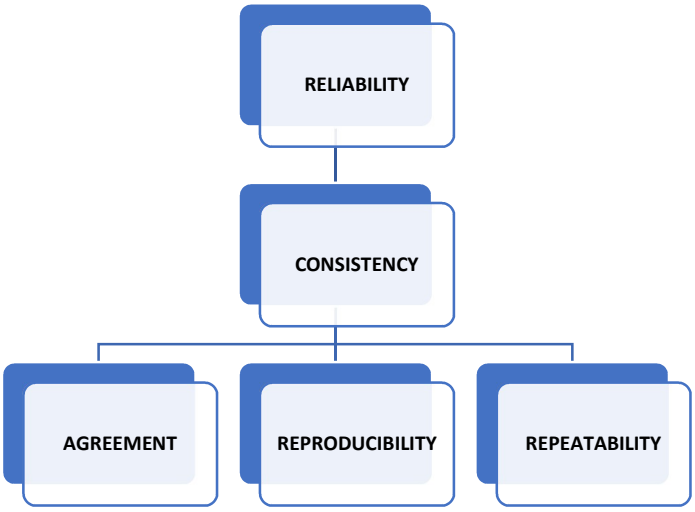

Figure 1. Reliability and its components

Validity simply means "measure what is intended to be measured" [13]. There are different types of validity in social sciences. Face validity is a subjective judgment on the operationalization of a construct

whether it is appropriate, unambiguous, simple, and proper [14]. Content validity refers to how appropriate and representative the measurements collected are for the desired assessment purpose. The representativeness criterion may have two definitions. Quantifying the extent of sampling is one of them. The second is the extent to which items reflect the structures of the whole scale [15]. Construct is a pattern formed by certain elements that are thought to be related to each other or by the relationships between them. The construct validity measurement tool shows to what extent it can accurately measure the structure and concept that it claims to measure [12]. Construct validity refers to how well you translated or transformed a concept, idea, or behavior that is a construct into a functioning and operating reality, the operationalization [14]. Construct validity is used when trying to quantify a hypothetical construct, like fear. Convergent and discriminant validity should be used to determine the validity of a construct by suggesting that the new measurements are correlated with other measurements of that construct and that the dimensions proposed are inappropriate to the construct unrelated, respectively [16]. Discriminant validity is the extent to which latent variable A discriminates from other latent variables. The Convergent Validity is the degree to which two measurements of a construct are connected theoretically [14]. The validity of the criterion is also divided into concurrent and predictive validity, where the validity of the criterion deals with the correlation between the current measurement and the criterion measurement (such as the gold standard) [16]. Content and construct validity in social sciences are defined as credibility / internal validity. Internal validity is related to the question of whether the research findings fit with reality in the external world. Internal validity is determined by experimenting with specific characteristics and no specific biases. For example, the question of "Can we recognize people by looking at their face?" can be examined. This question is answered by asking two more questions. First, is the independent variable the cause of the dependent variable? Second, can other possible explanations for the relationship between independent variable and dependent be logically eliminated? If the answer to these questions is yes, the researchers can claim that the experiment has internal validity [17]. Criterion validity is the degree to which it is empirically relevant to the outcome. This is something that calculates how well one measure predicts another measure. There are three types of criterion validity namely; concurrent validity, predictive and postdictive validity [14].

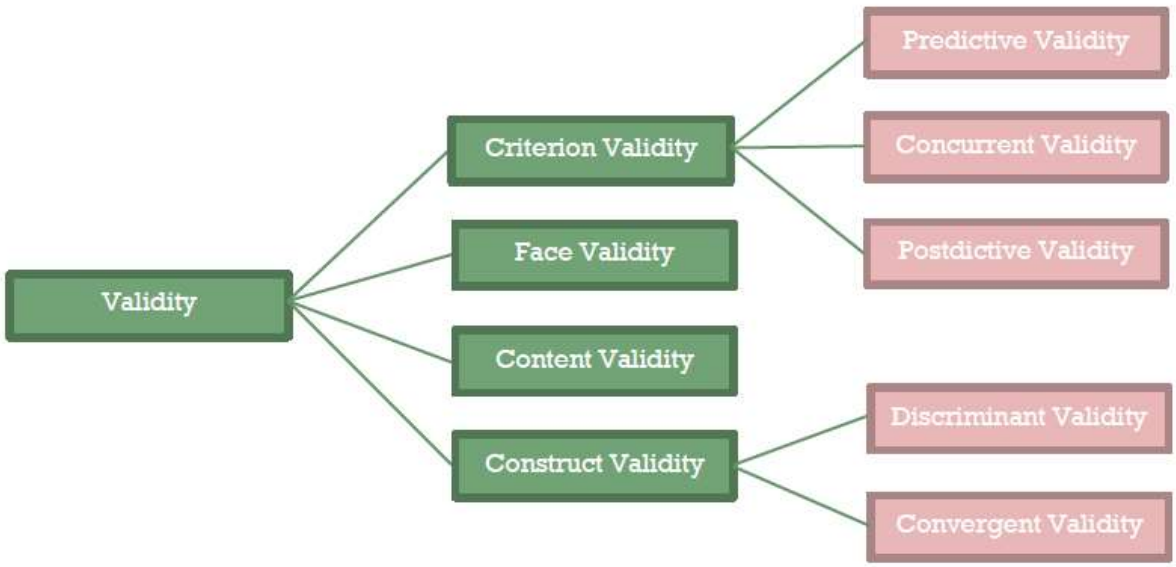

Figure 2. Subtypes of various forms of validity tests

Figure 1 illustrates how reliability and validity are related. In the first target, the shots reached the same spot, but none were effective in reaching the same point. The second target can be regarded as valid but not reliable since the points are expanding over the entire place. The third target did not present reliability or validity, since they hit spread points. The fourth target stands as an indicator of reliability and validity; the shots landed right in the target center and were consistent, right in the target center [18].

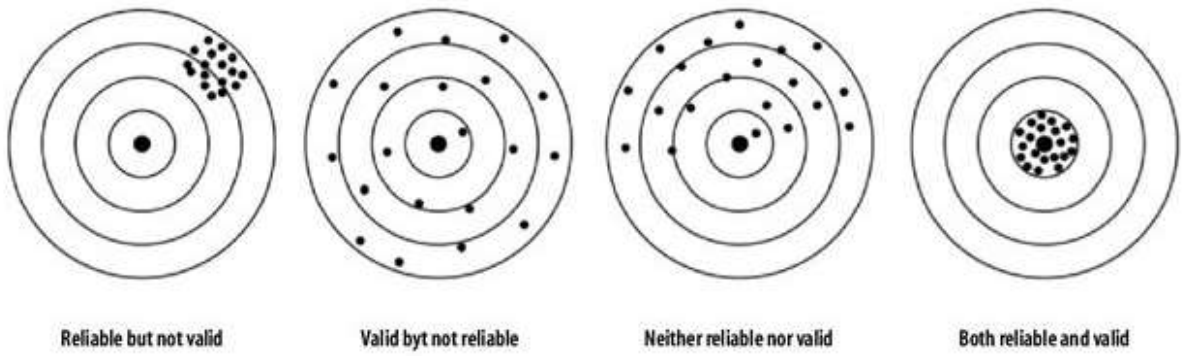

Figure 3. Possible combinations of validity and reliability of measurement instruments [18].

Sensitivity is defined as the consensus closeness between randomly selected individual measurements or results. It is therefore concerned with the variance of repeated measurements. A measurement tool with low variance is more sensitive than those with a higher variance. For example, as a researcher, one wants to know what is the smallest sample you can use that will take into consideration the variability in the dependent measure and yet be sensitive enough to notice a statistically meaningful difference, whether there is one. Our capacity to distinguish significant differences between groups is defined in part by the variability of individuals in our sample and how much variability occurs among them. Therefore, less variability may contribute to greater sensitivity, and more variability results in less sensitivity [19].

As mentioned above, the key aim of developing questionnaires or scales is to collect correct and appropriate data. The reliability and validity of scale or questionnaire formats is an important feature of testing methodology [14]. The reliable and accurate measurement may in the simplest intuitive terms indicate that the current measurement is equal with, or following, the truth. However, it is often impractical to require the new measurement to be identical to the truth, either because (1) we accept the measurement of a tolerable (or acceptable) error or (2) the truth is simply impossible for us (either because it is not measurable or because it is only measurable with some degree of error) [16]. In this regard data space and data range are the important dimensions of developing scales because it also changes the data type, the logical space of the analysis, methodology, and validity and reliability of the results.

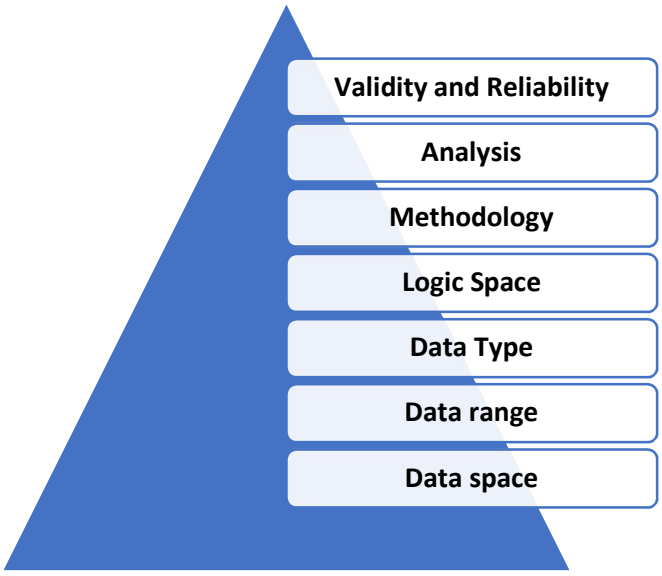

Figure 4. Data space and data range determines the validity and reliability of any scale

Data space in the measurement tools like scale refers to the set of the independent options regarding the particular item of the scale. For instance, in any Likert-type scale, there is only one option that can be expressed by the participant, therefore its data space is 1d, but in the neutrosophic scale, there are three independent dimensions regarding any item as undecided, agree, and disagree. As it can be seen data space is 1d in any Likert type scale and 3d in neutrosophic space and if our measurement tools become more qualitative like having items requiring free opinions in a paragraph like choices it has more dimensions even in ideal cases it has infinite dimensions. However, although n-dimensional space is more appropriate for better valid and reliable results, less dimensional spaces have less vagueness in terms of the interpretation of the data and they can be more easily statistically handled. Additionally, as the dimension of the space increases the objectivity of the measurement tool in terms of measuring common characteristics decreases. The advantage of the 3-dimensional neutrosophic scale is that it both seeks the agreement, disagreement, and confusion levels of the participants. In daily life, we encountered infinitely items to give an opinion about them and we are not restricted within a 1-dimensional space where we can only choose one answer regarding whether we agree, disagree express uncertain about a particular case. However, in the three-dimensional neutrosophic space participants express both their agreement and disagreement level as well as the uncertainty in the items or dimensions of the scale. We sometimes think that we understand a statement but one word in the statement makes us uncertain whether it is the "right meaning" intended by the source. Similarly, we sometimes agree on some propositions but just because of the source of the message itself we also disagree with the item. Therefore, the neutrosophic scale is different from the classical Likert-type scales in terms of data space.

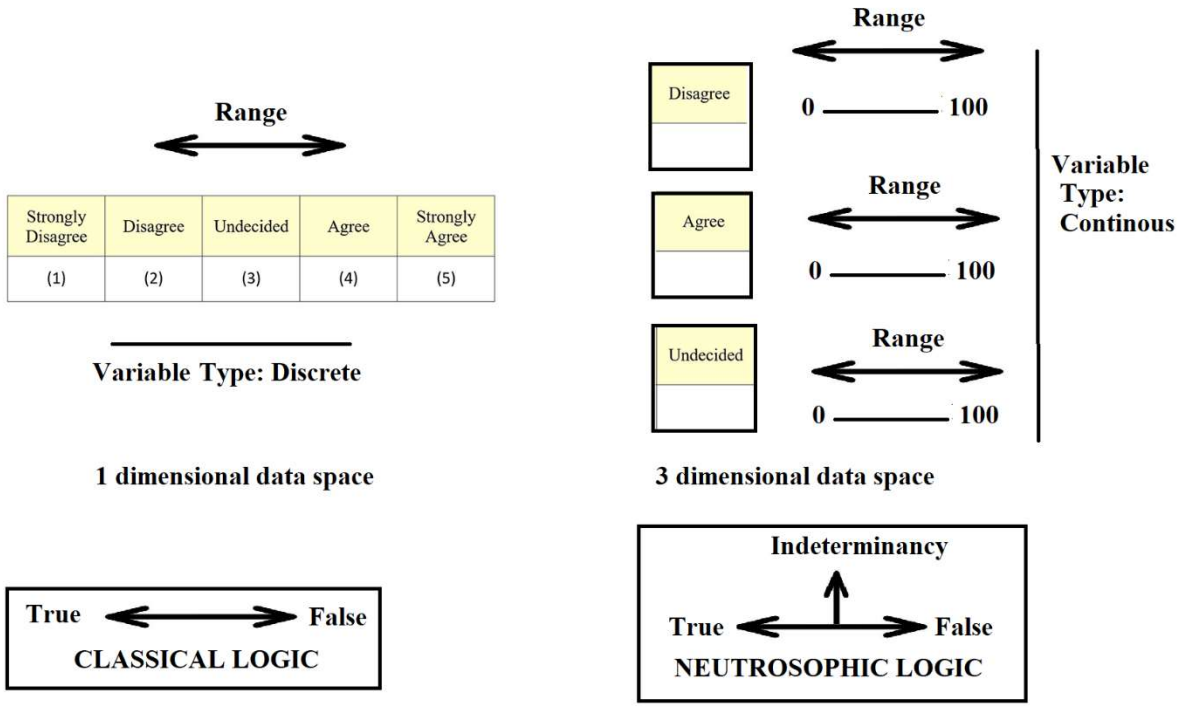

Figure 5. Data space of classical Likert type scale, neutrosophic scale

The second important point that distinguishes any measurement tool from each other is the data range. The range of a set of data is the difference between the highest and lowest values in the set. Likert type scales can be arranged in terms of data ranged commonly from 3 points Likert type to 10 point Likert type scales. However, the range of the neutrosophic scale is broader than the Likert type scales. It includes any rational number in a range between 0 and 100. Therefore, neutrosophic scales have continuous variables types whereas the Likert type scales have discrete value types in terms of rational numbers so that the analysis of the data may vary because of this. This can contribute to increasing the sensitivity of the measurement tool in this respect.

The third important point of any measurement tool is its logic space. The logic which is in our focus, Neutrosophic Logic, is an emerging field where each proposition is reckoned to have the proportion

(percentage) of truth in a subset T, the proportion of indeterminacy in a subset I, and the proportion of falsity in a subset F. We here consider a subset of truth (or falsity or indeterminacy), rather than just a number, since in many situations we do not have the capability to to precisely determine the proportions of truth and of falsity but we only have the ability to approach them. For example, suppose that a statement (or proposition) is between 32 % and 48 % true and 59 % to 73 % false; worse: 32 % to 39 % or 41 to 52 % true (according to various observers) and 57 % or 62 % to 71% false. Subsets are not basic intervals, but are any set (open or closed or semi-open/semi-closed intervals, discrete, continuous, intersections or unions of previous sets, etc.) in accordance with the given proposition. The adventure of gaining meaning and mathematical results from situations of uncertainty was initiated by Zadeh [20]. Fuzzy sets added a new wrinkle to the concept of classical set theory. Elements of the sets have degrees of belongingness (in another word, membership) according to underlying sets. Atanassov defined intuitionistic fuzzy sets including belongingness and non-belongingness degrees [21, 32, 33, 34]. Smarandache suggested Neutrosophy as a computational solution to the idea of neutrality [22]. Neutrosophic sets consider belongingness, non-belongingness and indeterminacy degrees. Intuitionistic fuzzy sets are defined by the degree of belongingness and non- belongingness and, uncertainty degrees by the 1-(membership degree plus non-membership degree), while the degree of uncertainty is assessed independently of the degree of belongingness and non- belongingness in neutrosophic sets. Here, belongingness, non- belongingness, and degree of uncertainty (uncertainty), like degrees of truth and falsity, can be assessed according to the interpretation of the places to be utilized. This indicates a difference between neutrosophic set and intuitionistic fuzzy set. The definition of neutrosophy is, in this sense, a potential solution and representation of problems in different fields. Two detailed and mathematical fundamental differences between relative truth (IFL) and absolute truth (NL) are as follows:

(i) NL has the capacity to distinguish absolute truth (truth in all possible worlds, according to Leibniz) from the relative truth (truth in at least one world) because NL (absolute truth) = 1^+ while IFL (relative truth) = 1. This has practice in philosophy and linguistics (see the Neutrosophy). The standard interval [0, 1] used in IFL has been extended to the unitary non-standard interval $]^- 0, 1^+[$ in NL. Parallel distinctiveness for absolute or relative falsehood and absolute or relative indeterminacy are allowed to consider in NL.

(ii) There do not exist any limits on T, I, F apart from they are subsets of $]^- 0, 1^+[$, thus: $^-0 \leq \inf T + \inf I + \inf F \leq \sup T + \sup I + \sup F \leq 3^+$ in NL. This permission allows dialetheist, paraconsistent, and incomplete information to be identified in NL, while these situations impossible to be identified in IFL since F (falsehood), T (truth), I (indeterminacy) are restricted either to $t + i + f = 1$ or to $t^2 + f^2 \leq 1$, if T, I, F are all reduced to the points t, i, f respectively, or to $\sup T + \sup I + \sup F = 1$ if T, I, F are subsets of [0, 1] in IFL.

Although in Likert-type scales, there are mostly three options as agreement, disagreement, and vagueness, its logic is located one valued option located on the opposite sides of true and false values. However, the neutrosophic set has three independent components giving more freedom for analysis so that it brings different logical operations as well. Therefore methodology of the analysis of the data should be changed based on the logical structure of the scale. For instance, while factor analysis is used for classical Likert type scales, neural networks are more appropriate for the analysis of the data of the neutrosophic scales. Nevertheless, it should be noted that classical analysis and methods can indeed be used for neutrosophic scales based on different analysis procedures. Hence, we can conclude that the validity, reliability of the measurement tools can change based on the logical structure of the scale. Therefore, in this study, we take The Satisfaction with Life Scale developed by Diener, Emmons, Larsen ve Griffin [23] and adapted in Turkish by Dağlı and Baysal [24] and convert this scale into neutrosophic form and compare the results of the scales and use this analysis for proposing new type confirmatory analysis procedure as well as developing neutrosophic scales.

1.2. The Difference between Lawshe Technique and Neutrosophic Scale

Some argue that well known Lawshe technique is very similar to neutrosophic analysis and propose what is the reason behind the logic of neutrosophic forms. Initially suggested in a seminal 1975 paper in Lawshe [25], the method of Lawshe was common in various areas including health care, education, organizational development, personnel psychology, and market research for determining and quantifying content validity [26,27]. First of all the main difference between those two techniques is in their data space. Although there are three choices in the Lawshe technique for each item as a- Essential? b-Useful but not essential? Why? c-Not necessary? Why which is very similar to the Truth T, indeterminacy I, and falsity F, regarding membership in neutrosophic logic, their dimensions are different from each other since there is only one option regarding for each item which corresponds to one-dimensional data space but there is three independent data space in the neutrosophic form where each data represents different. According to this, whether all participants agree that the information or ability that has been tested is necessary, or whether none says it is relevant, we are sure that the component has been added or omitted. If there is no majority, the dilemma emerges. There are two hypotheses, both compatible with existing psychophysical principles [28].

· Every item for which more than half of the experts consider any item to be "essential" has content validity.

· The wider the extent or degree of its validity is the more experts (above 50 percent) who view an item as "essential."

Therefore Lawshe technique focus on the dominant opinions of the experts which are restricted by one-dimensional data space so that it might hide their indeterminacy or disagree because they are weak comparing the other option. For a small number of items may be the effect of this can be negligible but for a huge number of items, it can make huge differences.

The second difference is related to the data range. Lawshe technique is restricted by the discrete data which can be manipulated by qualitative comments. Although qualitative comments make the item better, in terms of generalizability we may not be confident that the item is suitable for its content. Opinions of the experts may indicate different content but the understanding of common participants may indicate different content in this respect.

The third difference is related to statistical analysis. In the Lawshe technique, we focus on the ratio of decisions of the experts whereas in the neutrosophic logic we focus on the importance and correlation level of each item for the analysis. In the Lawshe technique, there is no distinction between the importance level and correlation so it leads that the item that is seen as important by experts might not be correlated with the content in the actual applications. In daily life, we wonder about the particular feature and we seek them in particular sets but the items of the set can be seen as important but are not relevant to what we want to seek. For example, we will meet a close relative whom we have not seen him/her for a long time and look for him/her in a particular place and the individuals looking like our relative are important for us but the importance will be diminished when we find that there is no correlation between the actual close relative and the similar person looking like him/her. Actually Satrre's [29] vivid description regarding his hypothetical appointment with Pierre can be given as more explicit example for the importance and correlation as follows:

> I have an appointment with Pierre at four o'clock. I arrive at the cafe a quarter of an hour late. Pierre is always punctual. Will he have waited for me? I look at the room, the patrons, and I say, "He is not here." Is there an intuition of Pierre's absence, or does negation indeed enter in only with judgment? At first sight it seems absurd to speak here of intuition since to be exact there could not be an intuition of nothing and since the absence of Pierre is this nothing…..
>
> Similarly Pierre's actual presence in a place which I do not know is also a plenitude of being. We seem to have found fullness everywhere. But we must observe that in perception there is always the construction of a figure on a ground. No one object, no group of objects is especially designed to be organized as specifically either ground or figure; all depends on the direction of my attention. When I enter this cafe to search for PIerre, there is formed a synthetic

organization of all the objects in the cafe, on the ground of which Pierre is given as about to appear. This organizatIon of the cafe as the ground is an original nihilation. Each element of the setting, a person, a table, a chair, attempts to isolate itself, to lift itself upon the ground constituted by the totality of the other objects, only to fall back Once more into the undifferentiation of this ground; it melts into the ground. For the ground is that which is seen only in addition, that which is the object of a purely marginal attention. Thus the original nihilation of all the figures which appear and are swallowed up in the total neutrality of a ground is the necessary condition for the appearance of the principle figure, which is here the person of Pierre. This nihilation is given to my intuition; I am witness to the successive disappearance of all the objects which I look at-in particular of the faces, which detain me for an instant (Could this be Pierre?) and which as quickly decompose precisely because they "are not" the face of Pierre. Nevertheless if I should finally discover Pierre, my intuition would be filled by a solid element, I should be suddenly arrested by his face and the whole cafe would organize itself around him as a discrete presence.

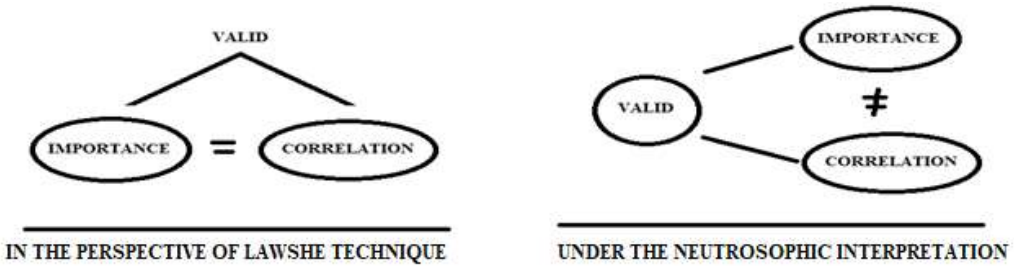

Figure 5. Importance and correlation are not equal in neutrosophic logic

Therefore, when experts make a decision there is no clear distinction for their decision-making process in terms of importance or correlation.

The fourth one is related to expert opinion. Lawshe technique focus on expert opinion but the term expert is not clear in many respects. For example, if somebody studies a novel concept that has not been studied previously how an expert decide whether the item is suitable or not except deciding on its grammar or meaning. Furthermore, we need different experts for decision making the suitability of the item but the ratio of those experts shouldn't be equal in the proportion of the decision-making process. For example in some scales the opinion of a psychologist might be more important than the other experts and their contribution should vary by this. However, in the neutrosophic scales, we mainly aim at the real participants so that we can understand to the extent whether the item is understood or objected or vague.

2.Methodology

In the methodology, firstly we convert the items of the Satisfaction with Life Scale into the neutrosophic form where each item has three independent components referring agreement, disagreement, and indeterminacy. However, to compare the neutrosophic scale we used also the classical scale as well. Secondly, we analyzed each item of neutrosophic scale in terms of classical scale in terms of neural networks and Spearman correlation constant. In the second part of the study, we analyzed the Neutrosophic Life Satisfaction Scale in terms of whole structure for confirmatory factor analysis. Finally, we created the decision-making formula to decide to remove or keep the items in the neutrosophic scale.

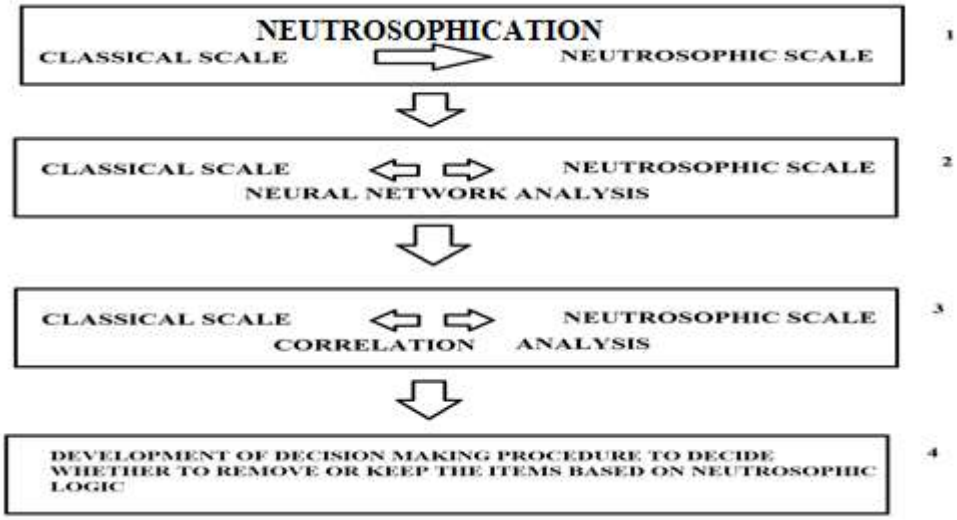

Figure 6. The procedure for the development of neutrosophic scale

In the neural network analysis for the study, for the level of the analysis of each item, the input variables are three sub-items of each item in the neutrosophic scale and the output variable is each classical scale. Similarly, for the whole structure for confirmatory factor analysis, the input variables are all the items in the neutrosophic scale and output variables are the classical items of the classical scale. Activation function both for hidden and output layer was chosen as sigmoid function. The number of hidden layers in each analysis were chosen to be two. Batch optimization was chosen to be criterıa trainİng=batch optimİzation=gradientdescent. In the analysis of the data independent variable importance analysis was used.

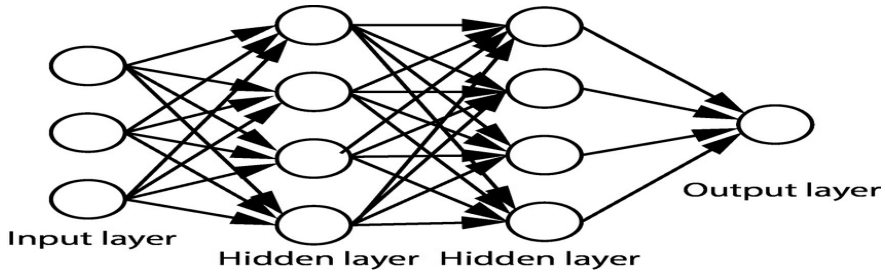

Figure 7. The general structure of the (CNN) convolutional neural network we used in this study as a three-layer neural network with three input neurons, two hidden layers of four neurons each, and one output layer [30]

"Independent variable importance analysis performs a sensitivity analysis, which computes the importance of each predictor in determining the neural network. The importance of an independent variable is a measure of how much the network's model-predicted value varies with different values of the independent variable. Normalized importance is just the importance values that are grouped by and represented as percentages of importance values. However, it should be underlined that you cannot tell is the "direction" of the relationship between these variables and the predicted probability of default" [31]. Therefore, we also look at the spearman correlations between the variables to see the direction and relationship of the items to decide whether they are suitable or not.

3. Findings

3.1. Analysis of Neutrosophic Life Satisfaction Scale in terms of Reliability

Before using the neutrosophic scale we can wonder its reliability before comparing it with the classical one. We can use Cronbach's Alpha constant for the neutrosophic scale. However, it should be noted that we should use Cronbach's Alpha constant three times for three independent factors as given in Table 3.1. below.

Table 3.1. Cronbach's Alpha constant for three dimensions

Cronbach's Alpha constant	Variables
,863	VAR1a VAR2a VAR3a VAR4a VAR5a
,777	VAR1b VAR2b VAR3b VAR4b VAR5b
,792	VAR1c VAR2c VAR3c VAR4c VAR5c

Results show that our neutrosophic scale is also reliable which also supports the reliability of the classical scale because Cronbach's Alpha constant is an acceptable level for three dimensions.

3.2.Analysis of Neutrosophic Life Satisfaction Scale in terms of items of validity

According to Spearman's rho correlation coefficient classical variable, 1 has a high positive significant correlation with var1a which is related to the agreeing level of the participants and it has an average level negative significant level of correlation variable 1c which is related to the disagreeing level of the participants. Both correlations can be related to the points of a participant who has either a high level of life satisfaction or not. Besides no correlation between vagueness and classical item shows that there is no indeterminacy about this item.

Table 3.2. Correlation among neutrosophic item 1 and classical item 1

		VAR1a	VAR1b	VAR1c
VAR1	Correlation Coefficient	,678**	-,022	-,417**
	Sig. (2-tailed)	,000	,768	,000
	N	189	189	189

Neural network analysis for the items shows that participants having positive life satisfaction for item 1a has 100% contribution to classical variable1 and participants having negative life satisfaction level for item 1c has 26,4% contribution to classical variable1. This might be related to the differentiation of the number of participants having high-level life satisfaction and a low level of life satisfaction. However, it should be noted that the vagueness of this item that is 57,5% implies that there is a moderate level of confusion about this article either because of meaning or the usage of the words or some unknown parameters, although there is no correlation between var1b and classical variable.

Table 3.3. Independent variable ımportance for classical item 1 in terms of neutrosophic items

Independent Variable Importance		
	Importance	Normalized Importance
VAR1a	,544	100,0%
VAR1b	,313	57,5%
VAR1c	,143	26,4%

According to Spearman's rho correlation coefficient classical variable2a has a significant positive correlation with var2a which is related to the agreeing level of the participants and it has variable2c has a negative significant low level of correlation which is related to the disagreeing level of the participants. Both correlations can be related to the points of participants who have either a high level of life satisfaction or not. Besides no correlation between vagueness and classical item shows that there is no indeterminacy about this item.

Table 3.4. Correlation among neutrosophic item 2 and classical item 2

		VAR2a	VAR2b	VAR2c
VAR2	Correlation Coefficient	,732**	,120	-,277**
	Sig. (2-tailed)	,000	,099	,000
	N	189	189	189

Neural network analysis for the items shows that participants having positive life satisfaction for item 2a has 100% contribution to classical variable2 and participants having negative life satisfaction level

for item 2c has 26,6% contribution to classical variable2. This might be related to the differentiation of the number of participants having high-level life satisfaction and a low level of life satisfaction. However, it should be noted that the vagueness of this item is 31,7% implies that there is a weak level of confusion about this article either because of meaning or the usage of the words or some unknown parameters, although there is no correlation between var1b and classical variable.

Table 3.5. Independent variable importance for classical item 2 in terms of neutrosophic items

Independent Variable Importance		
	Importance	Normalized Importance
VAR2a	,632	100,0%
VAR2b	,200	31,7%
VAR2c	,168	26,6%

According to Spearman's rho correlation coefficient classical variable, 3 has a moderate positive significant correlation with var3a which is related to the agreeing level of the participants and it has a negative significant moderate level of correlation which is related to the disagreeing level of the participants. Both correlations can be related to the points of participants who have either a high level of life satisfaction or not. However, the weak level of significant correlation between vagueness and classical item shows that there is an indeterminacy about this item.

Table 3.6. Correlation among neutrosophic item 3 and classical item 3

		VAR3a	VAR3b	VAR3c
VAR3	Correlation Coefficient	,474**	-,178*	-,430**
	Sig. (2-tailed)	,000	,014	,000
	N	189	189	189

Neural network analysis for the items shows that participants having positive life satisfaction for item 3a has 100% contribution to classical variable3 and participants having negative life satisfaction level fort he item 3c has 38, 0% contribution to classical variable3. This might be related to the differentiation of the number of participants having high-level life satisfaction and a low level of life satisfaction. However, it should be noted that the vagueness of this item 3c which is 21,7% implies that there is a weak level of confusion about this article either because of meaning or the usage of the words or some unknown parameters. It should be noted that there is also a weak level significant correlation with item 3b and item 3.

Table 3.7. Independent variable importance for classical item 3 in terms of neutrosophic items

Independent Variable Importance		
	Importance	Normalized Importance
VAR3a	,626	100,0%
VAR3b	,136	21,7%
VAR3c	,238	38,0%

According to Spearman's rho correlation coefficient classical variable, 4 has a high-level significant correlation with var4a which is related to agreeing on the level of the participants and it has a negative moderate level significant correlation which is related to the disagreeing level of the participants. Both correlations can be related to the points of participants who have either a high level of life satisfaction or not. Besides no correlation between vagueness and classical item shows that there is no indeterminacy about this item.

Table 3.8. Correlation among neutrosophic item 4 and classical item 4

		VAR4a	VAR4b	VAR4c
VAR4	Correlation Coefficient	,715**	-,115	-,475**
	Sig. (2-tailed)	,000	,115	,000
	N	189	189	189

Neural network analysis for the items shows that participants having positive life satisfaction for item 4a has 95,8 % contribution to classical variable4 and participants having negative life satisfaction level fort he item 4c has 100, 0% contribution to classical variable4. This might be related to the differentiation of the number of participants having high-level life satisfaction and a low level of life satisfaction. However, it should be noted that the vagueness of this item 4c that is 27,0% implies that there is a weak level of confusion about this article either because of meaning or the usage of the words or some unknown parameters, although there is no correlation between variable 4b and classical variable.

Table 3.9. Independent variable importance for classical item 3 in terms of neutrosophic items

	Independent Variable Importance	
	Importance	Normalized Importance
VAR4a	,430	95,8%
VAR4b	,121	27,0%
VAR4c	,449	100,0%

According to Spearman's rho correlation coefficient classical variable, 5 has a high level of positive significant correlation with var5a which is related to the agreeing level of the participants and it has a weak level of negative significant correlation which is related to the disagreeing level of the participants. Both correlations can be related to the points of participants who have either a high level of life satisfaction or not. Besides, there is a weak level significant correlation between variable 5 and variable 5b. Therefore, the weak level significant correlation between vagueness and classical item shows that there is an indeterminacy about this item.

Table 3.10. Correlation among neutrosophic item 5 and classical item 5

		VAR5a	VAR5b	VAR5c
VAR5	Correlation Coefficient	,706**	,149*	-,347**
	Sig. (2-tailed)	,000	,040	,000
	N	189	189	189

Neural network analysis for the items shows that participants having positive life satisfaction for item 5a has 100,0 % contribution to classical variable4and participants having negative life satisfaction level fort he item 5c has 84, 2 % contribution to classical variable4. This might be related to the differentiation of the number of participants having high-level life satisfaction and a low level of life satisfaction. However, it should be noted that the vagueness of this item 4c that is 39,6% implies that there is a weak level of confusion about this article either because of the meaning of the usage of the words or some unknown parameters.

Table 3.11. Correlation among neutrosophic item 5 and classical item 5

	Independent Variable Importance	
	Importance	Normalized Importance
VAR5a	,447	100,0%
VAR5b	,177	39,6%
VAR5c	,376	84,2%

3.2. Analysis of Neutrosophic Life Satisfaction Scale in terms of whole structure for confirmatory factor analysis

Neural network analysis results for two scales can be given as follows. It seems that variable 2 and variable 5 might be problematic when considering the overall contribution of the items for the whole scale since variable …b items are related to the vagueness of the participants.

Table 3.12. Independent variable importance for the whole scales

Independent Variable Importance		
	Importance	Normalized Importance
VAR5c	0,162	100,00%

VAR2a	0,133	82,30%
VAR5a	0,121	74,70%
VAR3a	0,1	61,50%
VAR1c	0,096	59,30%
VAR2b	0,09	55,70%
VAR5b	0,083	51,10%
VAR4a	0,075	46,60%
VAR3c	0,035	21,50%
VAR1a	0,032	20,00%
VAR2c	0,022	13,30%
VAR4b	0,018	11,20%
VAR4c	0,015	9,00%
VAR1b	0,013	7,80%
VAR3b	0,005	2,90%

4. Discussion and Conclusion

Content validity refers to how appropriate and representative the measurements collected are for the desired assessment purpose. Content validity refers to how appropriate and representative the measurements obtained are for the desired assessment purpose. The representativeness criterion may have two definitions. Quantifying the extent of sampling is one of them. The second is the extent to which items reflect the structures of the whole scale [15]. In this respect, the most eliminating factor to decide whether the item should be removed or not is to use the vagueness choices of the participants for each item. In this respect, we have two kinds of variables to formalize our decision making as correlation constant and importance level. If we label the decision function as d where r stands for correlation constant and I stands for importance level, the function for decision making can be written as like this:

$$D=R*I \tag{4.1}$$

The interpretation of this formula can be given in Table 4.1. It should be noted that the correlation constant is the absolute value of r as $|R|$.

Table 4.1. The interpretation of the formula D=R*I

The interpretation of the correlation coefficient (r);	**The interpretation of the importance level**	**Decision Criteria for accepting or rejecting the item where 0<cc<1** **Decision=[correlation coefficient for vagueness (r)]*[Importance level for vagueness]**
Very weak correlation or no correlation if r <0.2	Very weak importance level if <20%	if 0≤cc≤20, item acceptable
Weak correlation between 0.2-0.4	Weak importance level 20%-40%	if 20<cc≤40, item acceptable
A moderate correlation between 0.4-0.6	Moderate importance level 40%-60%	if 40<cc≤60, the item should be modified or removed
The high correlation between 0.6-0.8	High importance level 60%-80%	if 60<cc≤80, the item should be modified or removed
If r>0.8, it is interpreted that there is a very high correlation	If 80%>, it is interpreted that there is a very high importance level	if 80<cc≤100, the item should be removed

We can apply the formula 4.1 for the findings of the items of the Neutrosophic Life Satisfaction Scale for confirmatory analysis. Let's look at our findings based on item levels with the formula (4.1) as given in Table 4.2. The results show that this scale is valid because all the items are at an acceptable level.

Table 4.2. Application of the formula (4.1) for each item

	Importance Level (i)	Correlation constant (r)	Decision result (d=i*r)	
Var1	57,5	0,22	12,65	Acceptable
Var2	31,7	0,12	3,804	Acceptable
Var3	21,7	0,178	3,8626	Acceptable
Var4	27	0,115	3,105	Acceptable
Var5	39	0,149	5,811	Acceptable

In Table 3.11, independent variable importance for the whole scale shows that variable 2 and variable 5 might be problematic when considering the overall contribution of the items for the whole scale since variable …b items are related to the vagueness of the participants. However the formula 4.1 shows that although the importance level them are high they are not significant so that all the items of the scale are valid. Finally one might ask that if we only focus on the item related to vagueness why we need the other two items regarding agreement and disagreement? Although in this scale we don't see such a conflict, we can use those data to evaluate the validity and reliability of the scale. For instance, if both agreement and disagreement items have a similar sign with the target item, we can conclude that this item is also problematic because it reflects both agreement and disagreement at the same time implying that there is confusion about it for determining the aimed question. Let label that the correlation of agreement item is α and the correlation of disagreement item is β since these items are opposite to each other their correlation should naturally be opposite to each other so that $\alpha*\beta=-1$. If $\alpha*\beta=+1$ we can conclude that there is a contradiction in this item. If we modify equation 4.1 for these values where i_1 is the importance level of the first item and i_2 is the importance level of the second item as follows

$$(i_1 * \alpha * i_2 * \beta)/100 = d \qquad (4.2)$$

We divide 100 the multiplication just because scaling the value in a more simple form because we don't want to deal with huge numbers in case all the importance level 100 and correlations 1 or -1. Let us apply the rule of our correlation constants in the finding section for each item in Table 4.3. An opposite sign indicates that our data is consistent, otherwise, we should look at the effect of the correlations and evaluate whether we remove the item or not just as in the classification given in Table 4.1.

Table 4.3. Decision matrix evaluating the consistency of the items in terms of agreement and disagreement items of the neutrosophic scale

	i_1	A	i_2	β	$i_1 * \alpha * i_2 * \beta$	Decision
Variable 1	100	0,678	26,4	-0,417	-7,4639664	Acceptable
Variable 2	100	0,732	26,6	-0,277	-5,3935224	Acceptable
Variable 3	100	0,474	38	-0,430	-7,74516	Acceptable
Variable 4	95,8	0,715	100	-0,475	-32,536075	Acceptable
Variable 5	100	0,706	84,2	-0,347	-20,6274844	Acceptable

5. Future Directions

We can use a neutrosophic scale to confirm the reliability of the classical one because the neutrosophic scale is just an extended form of the classical one. Our results show that our neutrosophic scale is also reliable which also supports the reliability of the classical scale because Cronbach's Alpha constant is an acceptable level for three dimensions. In this respect, we can understand the Agreement dimension of reliability because we extend the classical scale into the neutrosophic one and assess the closeness of two measurements made on the same subject as opposed to one another. We also assess the Repeatability of the scale because we measure the same variable again and again for the same circumstances [9]. We

also test the Reproducibility of the scale because we can test the variations in test results while tests were performed on subjects on different occasions.
Validity simply means "measure what is intended to be measured" [13].To decide whether a scale is valid or not we can compare its validity by comparing similar scales or we make decisions based on expert opinion. In this study, we offer an alternative method for developing a valid scale where first we convert the scale into a neutrosophic one and then compare them through neural networks. We can understand also any scale to assess how appropriate and representative the measurements collected are for the desired assessment purpose so that we can evaluate its content validity. We can understand how well we translated or transformed a concept, idea, or behavior that is a construct into a functioning and operating reality, the operationalization [14] in any scale so that we understand its construct validity. We can use this method for also criterion validity because we can calculate how well one measure predicts another measure.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare that they have no conflict of interest.

Authorship Contributions: The authors declare that they contribute equally to the study.

6. References

[1] Eliot J. (1987) *The Psychometric Approach. In: Models of Psychological Space*. Springer, New York, NY. https://doi.org/10.1007/978-1-4612-4788-3_3.

[2] Özgüven, İ. E. (2011). *Psikolojik testler*. Ankara: Pdrem Yayınları.

[3] Atılgan, H. (2009). *Eğitimde ölçme ve değerlendirme*. Ankara: Anı Yayıncılık.

[4] Fraenkel, J.R. Wallen, N. E., Hyun, H.H. (2012*). How to Design and Evaluate Research in Education*, USA: The McGraw-Hill Companies.

[5] Tavşancıl, E. (2010). *Tutumların ölçülmesi ve SPSS ile veri analizi*. Ankara: Nobel Yayıncılık.

[6] Hinkin, T. (1995). A Review of Scale Development Practices in The Study of Organizations. *Journal of Management*, 21(5), 967-988.

[7] Şencan, H. (2005). *Sosyal ve Davranışsal Ölçümlerde Güvenilirlik ve Geçerlilik*. Ankara: Seçkin Yayıncılık.

[8] Carmınes, E. G. & Zeller, R. A. (1979). *Reliability And Validity Assessment*, Newbury Park, Ca, Sage.

[9] Bartlett JW, Frost C. (2008). Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound Obstet Gynecol*. Apr;31(4):466-75. doi: 10.1002/uog.5256. PMID: 18306169.

[10] Moser, C. A. & Kalton, G. (1989). *Survey Methods In Social Investigation*, Aldershot, Gower.

[11] Rosen, A.M. (2019). *Effective Research Methods for Any Project*, USA: Teaching Company.

[12] Hergüner, S. (2010). *Ölçme Araçlarının Kullanımı İle İlgili Temel Kavramlar, Çocuk Ve Ergen Psikiyatrisinde Ölçüler ve Ölçekler*, Türkiye Çocuk ve Genç Psikiyatrisi Derneği, İstanbul

[13] Fıeld, A. P. 2005. *Discovering Statistics* Using SPSS, Sage Publications Inc

[14] Taherdoost, H. (2016). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research, *International Journal of Academic Research in Management (IJARM)*, 5(3), 28-36

[15] Slavec, A. & Drnovsek, M. (2012). Perspective in scale development in entrepreneurship research. *Economic and Business Review*. 14 (1), 39-62.

[16] Barnhart, H. & Haber, M. & Lin, L. (2007). An Overview on Assessing Agreement with Continuous Measurements. *Journal of biopharmaceutical statistics*. 17. 529-69. 10.1080/10543400701376480.

[17] Vanderstoep, S.W., and Johnston, D.D. (2009) *Research Methods for Everyday Life Blending Qualitative and Quantitative Approaches*. Jossey-Bass, San Francisco.

[18] De Souza, A. & Alexandre, N. & de Brito Guirardello, E. (2017). Psychometric properties in instruments evaluation of reliability and validity. *Epidemiologia e Serviços de Saúde*. 26. 649-659. 10.5123/S1679-49742017000300022.

[19] Mertens,D.M.(2010). *Research and evaluation in education and psychology: integrating diversity with quantitative, qualitative, and mixed methods*, USA: Sage.

[20] Zadeh, L.A. (1965). Fuzzy sets. *Inf. Control,* 8, 338–353.

[21] Atanassov, K.T. (1986). Intuitionistic fuzzy sets. *Fuzzy Sets Syst*. 20, 87–96.

[22] Smarandache, F. (1998). *Neutrosophy. Neutrosophic Probability, Set, and Logic*; American Research Press: Rehoboth, DE, USA; 104–106.

[23] Diener, E., Emmons, R. A., Larsen, R. J., and Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, 49 (1), 71-75.

[24] Dağlı, A. ve Baysal, N. (2016). Yaşam Doyumu Ölçeğinin Türkçe'ye uyarlanması: Geçerlik ve güvenirlik çalışması. *Elektronik Sosyal Bilimler Dergisi* , 15(59), 1250-1262.

[25] Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28, 563–575.

[26] 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, 197–210. doi:10.1177/0748175612440286

[27] Ayre, C. & Scally, A. (2013). Critical Values for Lawshe's Content Validity Ratio. *Measurement and Evaluation in Counseling and Development.* 47. 79-86. 10.1177/0748175613513808.

[28] Gilbert, G. & Prion, S. (2016). Making Sense of Methods and Measurement: Lawshe's Content Validity Index. *Clinical Simulation in Nursing*. 12. 530-531. 10.1016/j.ecns.2016.08.002.

[29] Sartre, Jean-Paul (2003). Being and Nothingness. Hazel E. Barnes (trans.). London: Routledge. pp. 9-10.

[30] Nait Aicha, A., Englebienne, G., van Schooten, K. S., Pijnappels, M., & Kröse, B. (2018). Deep Learning to Predict Falls in Older Adults Based on Daily-Life Trunk Accelerometry. *Sensors* (Basel, Switzerland), 18(5), 1654. https://doi.org/10.3390/s18051654

[31] IBM SPSS Neural Networks 21 (2011). *IBM SPSS Neural Networks 21 Manual*, USA: IBM,pp.53-54

[32] Çevik, A. Topal, S. Smarandache, F. (2018). Neutrosophic Logic Based Quantum Computing, *Symmetry*, 10, 656

[33] Smarandache, F. (1998). *Neutrosophy: Neutrosophic Probability, Set and Logic*; American Research Press: Rehoboth, NM, USA.

[34] Taş, F., Topal, S., & Smarandache, F. (2018). Clustering neutrosophic data sets and neutrosophic valued metric spaces. Symmetry, 10(10), 430.

Figures

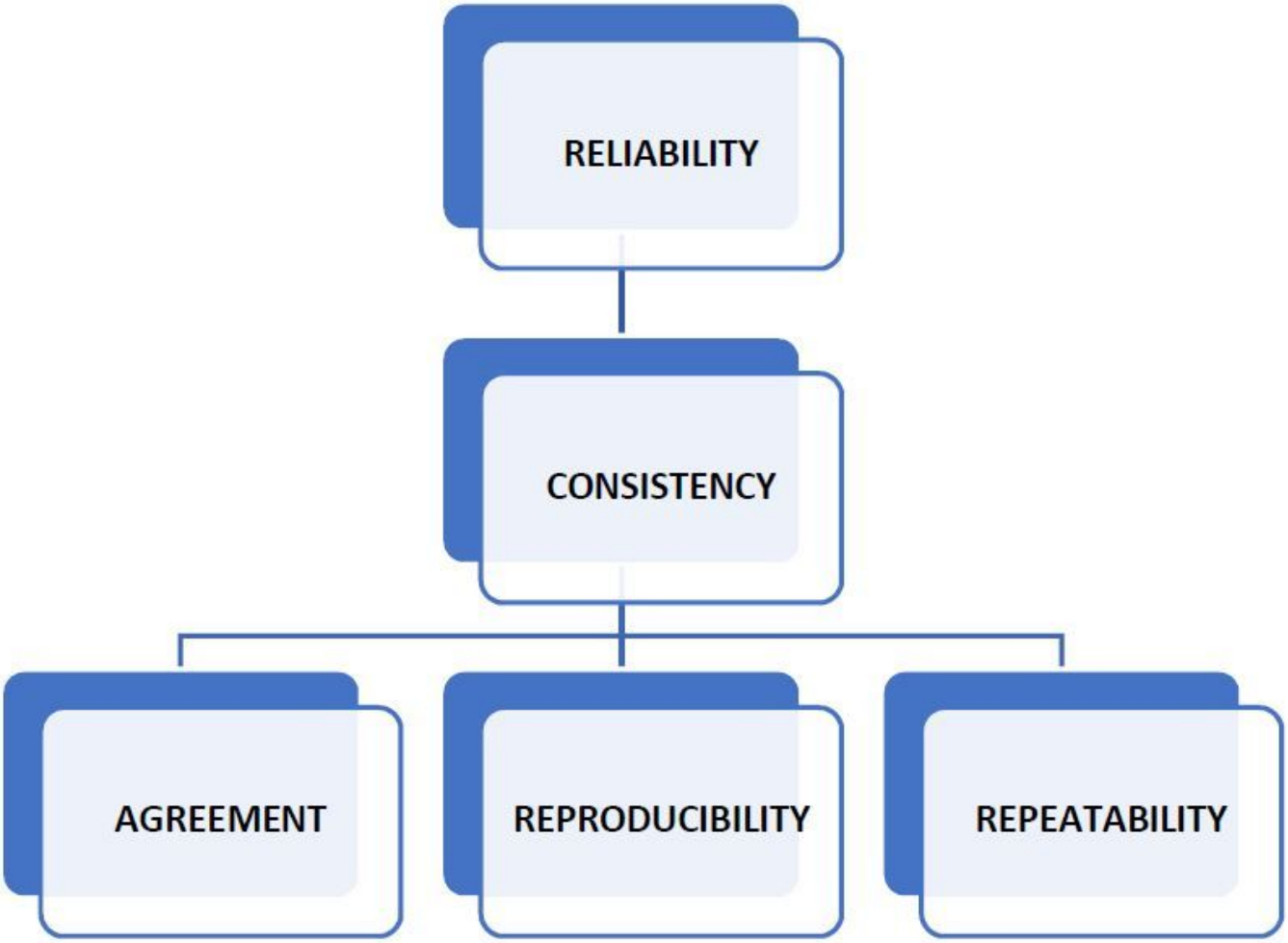

Figure 1

Reliability and its components

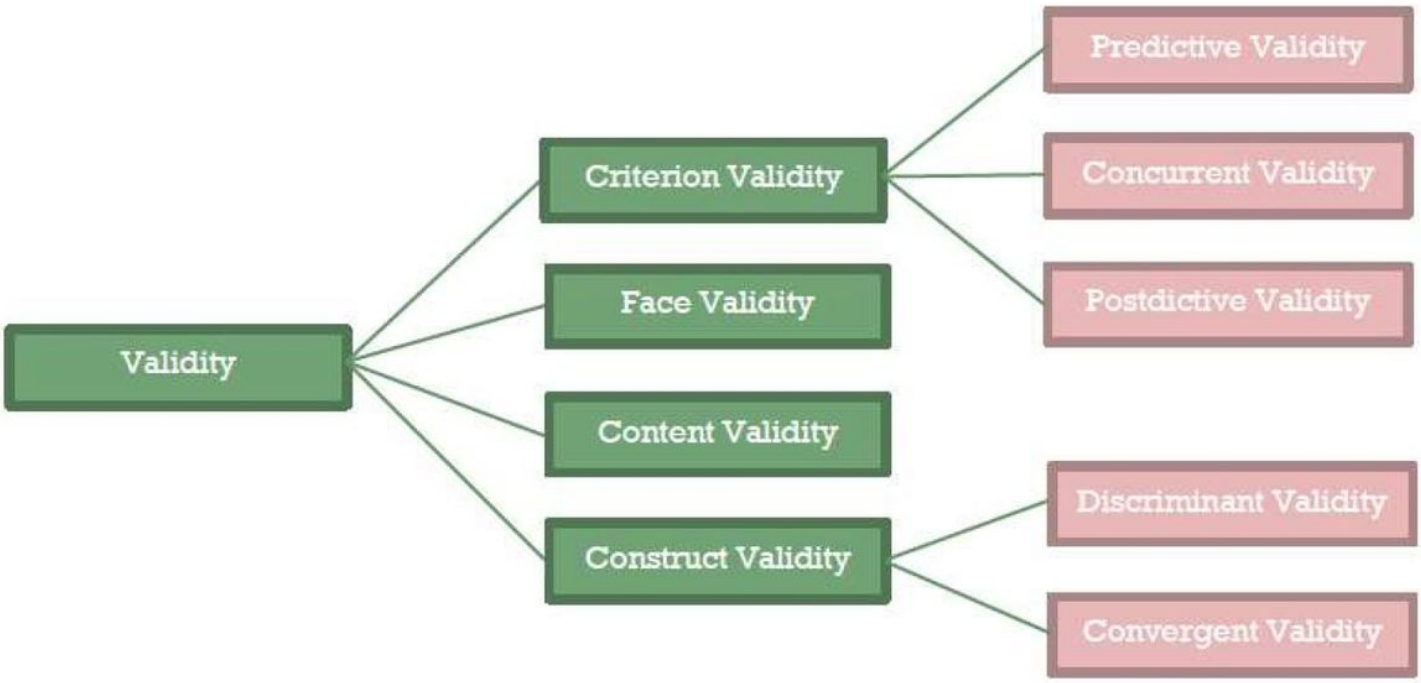

Figure 2

Subtypes of various forms of validity tests

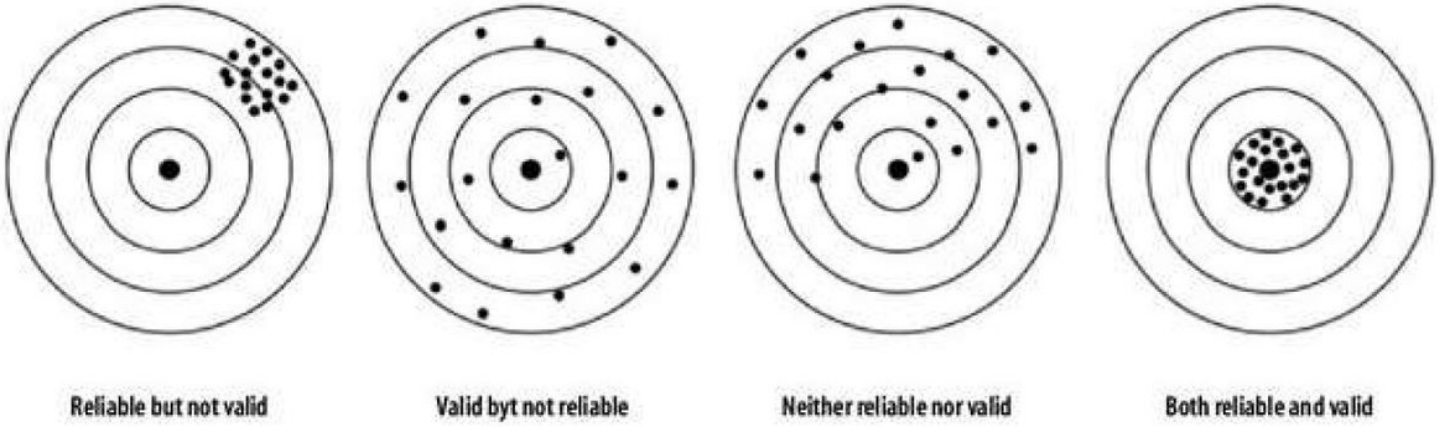

Figure 3

Possible combinations of validity and reliability of measurement instruments [18].

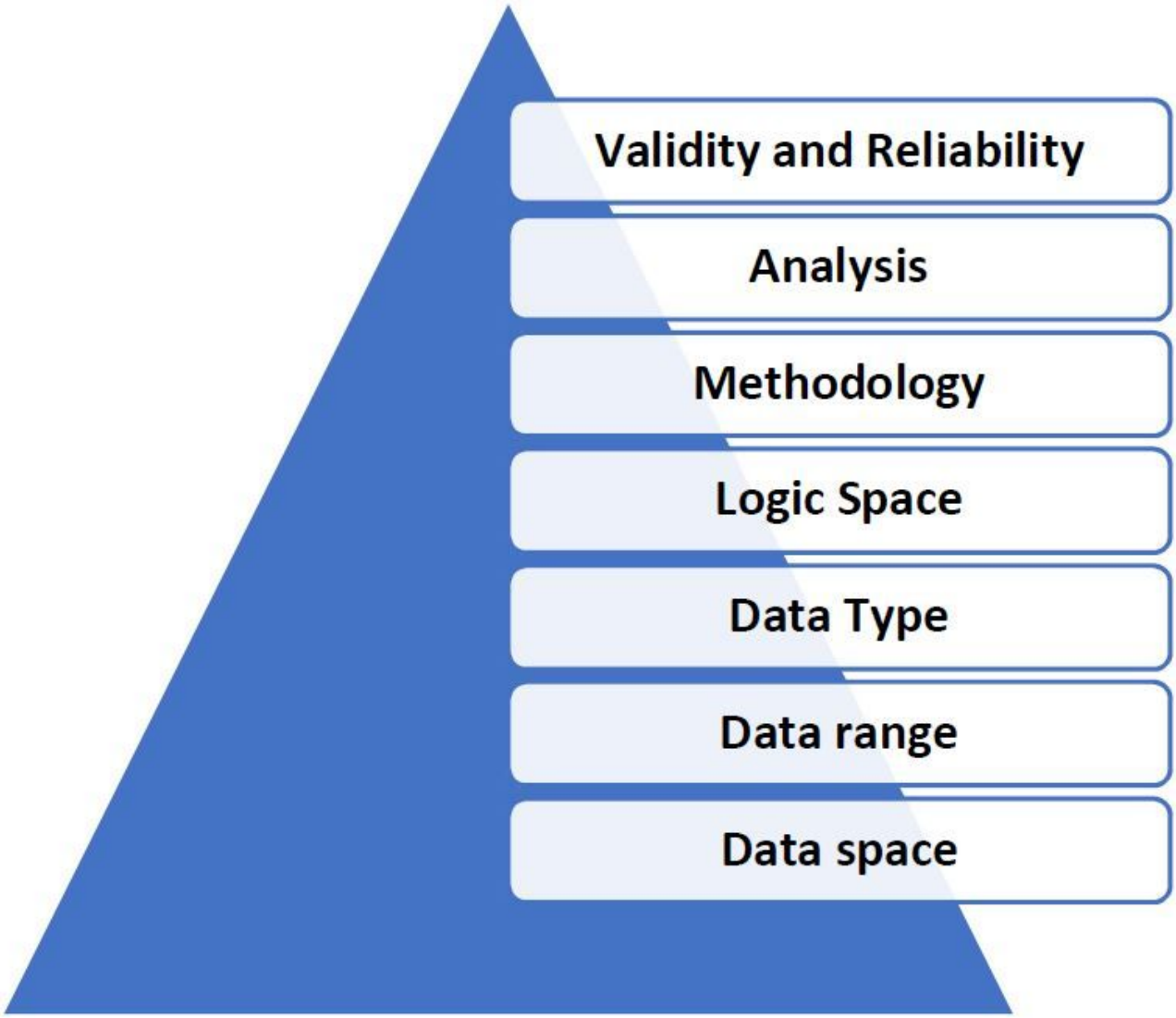

Figure 4

Data space and data range determines the validity and reliability of any scale

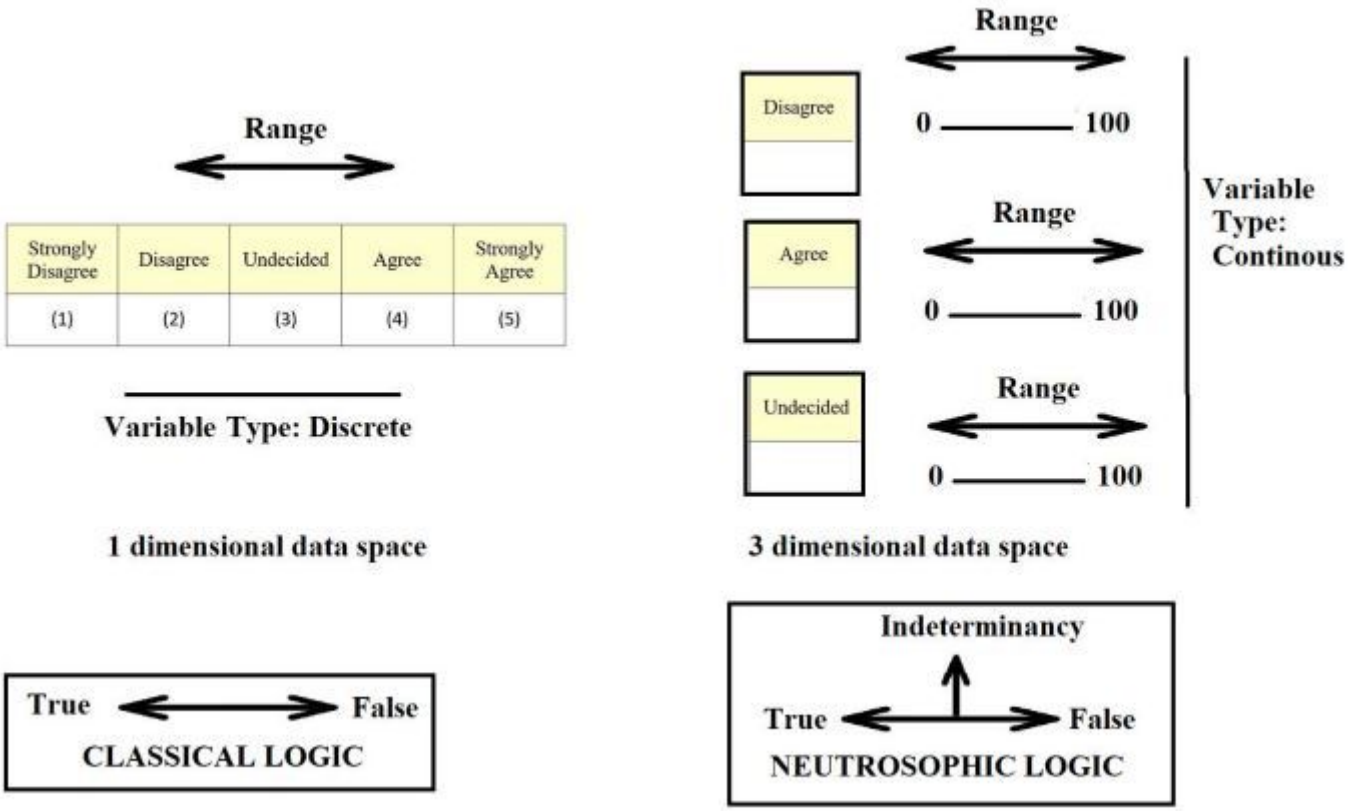

Figure 5

Data space of classical Likert type scale, neutrosophic scale

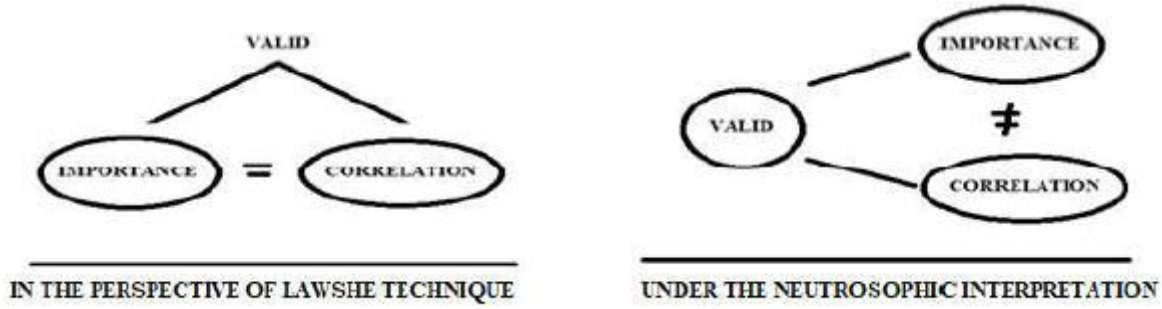

Figure 6

Importance and correlation are not equal in neutrosophic logic

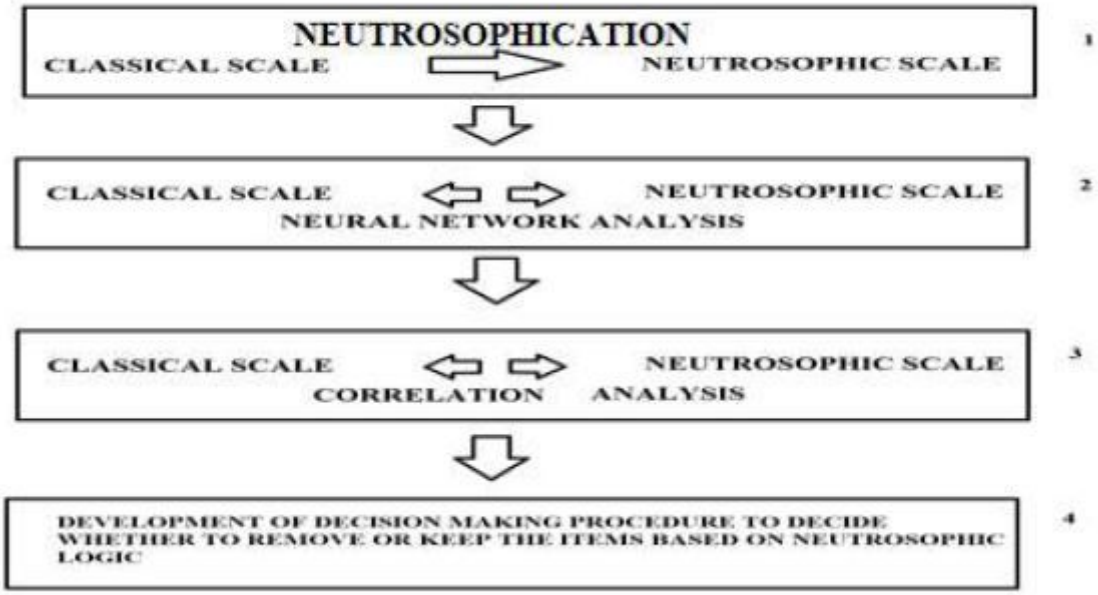

Figure 7

The procedure for the development of neutrosophic scale

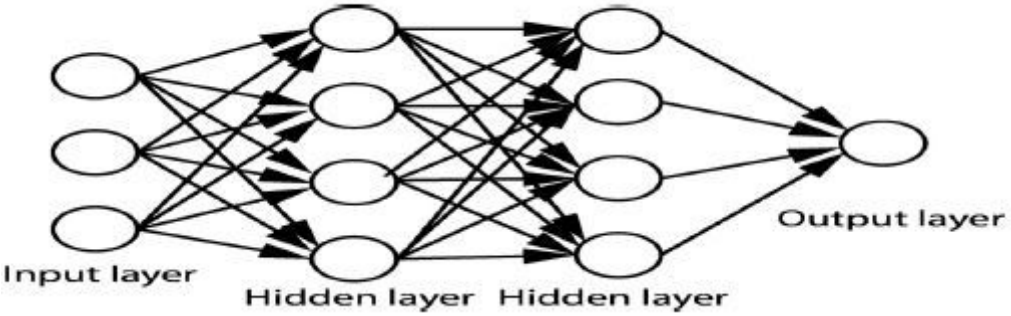

Figure 8

The general structure of the (CNN) convolutional neural network we used in this study as a three-layer neural network with three input neurons, two hidden layers of four neurons each, and one output layer [30]