Building a Self-Confidence Scale According to the Item Response Theory for High School Students in Jordan

Hani Alkhaldi[1], Malek Alkhutaba[2] & Mohammad Al-Dlalah[1]

[1] Measurement and Evaluation, Department of Teacher Education-Isra University, Jordan

[2] Department of Psychology-Isra University, Jordan

Correspondence: Hani Alkhaldi, Measurement and Evaluation, Department of Teacher Education-Isra University, Jordan.

Abstract

This study aimed to build self-confidence for high school students in Al-Mafraq Governorate in Jordan following the Item Response Theory (IRT). The scale included its initial version (50) items. To ensure the external validity of the scale, it was reviewed by several experts. According to the experts' feedback, some items should be deleted or modified. The final version of the scale included (44) items. The scale was further applied to an experimental sample of (310) male and female students to verify psychometricians' characteristics. Finally, the scale was administered to a sample of (1060) male and female high school students in Al-Mafraq Governorate. Data were collected, coded, and analyzed using statistical programs (SPSS and WINSTEPS). The most important results were the following: the self-confidence measure was one-dimensional, which means it measures only a single dimension. The results further revealed identical to the partial estimation model, and the index of average matching of individuals and the external and internal items approached zero, and the standard deviation approached the correct one. The estimated values of the distinct thresholds for the scale items showed a clear discriminatory ability and the emergence of particular threshold scores on the scale. After deleting the paragraphs that did not fit the study's model, the scale's final version included 39 items. The results also showed that the transfer values of logistical capacity units were within (-2.88 -2.77), within the IRT's accepted range.

Keywords: Modern Measurement Theory, differential thresholds, logistical capacity, one-dimensional, Item Response Theory (IRT)

1. Introduction

The developed and developing countries of the world pay great attention to the education process in all stages. As they progress, the development and prosperity of any society in various fields of life depends to a large extent on education and scientific research; the educational process helps in developing people and exploding their scientific capabilities, building their personality, and making them capable of confronting problems and solving them scientifically. Therefore, a person who has self-confidence is more concerned; has more desire to give; appreciate time; is keen to solve any problem that he may face; and tends to challenge any obstacles (Al-Rikabi, 2000). Moreover, self-confidence makes people able to get rid of any anxiety they might face in life. A self-confident person is aware of his abilities and potentials for life and study because he knows himself, indicating his compatibility and good mental health (Al-Badrani, 1986).

Many studies (e.g., Bunker, 1991; Nassif, 1993) have emphasized the importance of building and developing self-confidence since childhood. Furthermore, some other studies (e.g., Gesit & Hamrick, 1983; Al-Anzi, 2001) have found a positive relationship between self-confidence and good mental health. In contrast, other studies have found the compatibility of self-confidence with mental health, realistic thinking, sense of competency, vitality, activity, and the ability to withstand crises (Barron, 1953; Kleirmuntz, 1960; Cottel, 1974). Therefore, the importance of self-confidence is evident for high school students, especially that self-confidence is related to a positive and statistically significant relationship with achievement, as confirmed by suggested by (Abu Alam, 1978; Fennema & Sherman, 1978). In addition to Abu Allam (1978), who designated that low self-confidence is one of the most serious problems that high school students suffer from.

Building measures need to adopt a specific theory to verify the scale's psychometric properties' quality in light of its assumptions. Therefore, this study counted on Item Response Theory (IRT) to respond to the classical theory's shortcomings. Furthermore, IRT provided more accurate methods in estimating the degree of possession of the characteristic under investigation and item parameters (difficulty, discrimination, prediction, information function). Because of the lack of standard scales in Jordan, according to our knowledge, and due to the urgent need for such measures, the present study was initiated to build a scale of self-confidence among high school students.

2. The Problem of the Study

Despite the technological progress, high school students still face many social and psychological problems. Perhaps the most prominent of these problems is the lack of self-confidence. Previous studies suggested a lack of self-confidence among secondary school students, especially females, which resulted in poor achievement and a negative outlook for the future (Ali, 2009). It was found that secondary school students' self-confidence stimulated their behavior towards pursuing oneself by achieving a high level of self-confidence and achieving high academic levels (Kawthar, 2007). Likewise, positive self-beliefs increase students' self-confidence, which drives them to achieve excellence and success (Abu El-Cell, 2016).

One of the characteristics of self-confidence is that it stimulates positive emotions, creates a feeling of enthusiasm and joy, helps focus attention, increases perseverance and effort to achieve goals and success. Consequently, this can contribute to building a positive concept, making the individual feel more comfortable and freer of fears, organizing his thoughts quickly and accurately, and less help from others. It thus enables him to overcome difficulties and reach a high level of academic achievement, which may lead to self-respect and eagerness to engage with discussions with others (Al-Omar, 2000). Hence, this motivated the researchers to build a scale of self-confidence for secondary school students in Jordan according to the Item Response Theory (IRT).

3. Research Questions

Q1: What are the one-dimensional implications of the items of the self-confidence scale?

Q2: What are the implications of matching the self-confidence scale's empirical data for the one-dimensional partial estimate model?

Q3: What are the estimates of the distinctive threshold values of the items' self-confidence scale?

Q4: What are the performance standards for high school students on the scale of self-confidence in terms of transferred logistical units?

Objectives of the study

The current study aimed to build a measure of self-confidence for high school students in Jordan using the modern theory of measurement as a reference for measuring the self-confidence of high school students.

Significance of the study

In theory, this study provides facts, information, different references, results, and studies on the trait of self-confidence. It is argued that most of the individual's positive personality aspects (*e.g.,* independence, self-realization, ambition, and achievement) do not grow except with self-confidence growth. Therefore, the current study is an attempt to build a scale of self-confidence among high school students. In practice, the study's importance lies in creating a scale of self-confidence to identify its psychometric properties according to the IRT.

Terminology

Self-confidence scale: It is a set of stimuli prepared in verbal positions to quantitatively measure high school students' self-confidence in Al-Mafraq Governorate in Jordan. Self-confidence is defined idiomatically: "It is the individual's belief in his abilities to run his affairs without fear, to achieve his goals, his self-acceptance as it is, and his belief that he is worthy of the esteem of others" (Al-Rifai, 2004). It is also known as: "A measure of an individual's perception of himself and his belief in his abilities that depend on his previous experiences" (Perry, 2011).

Self-confidence is defined idiomatically procedurally: as "the total score obtained by the respondent through his response to the items of the scale that was built in this study."

Estimation of Ability:

It indicates how much of an attribute an individual possesses and can be identified with a point on the attribute continuum ($j\varnothing$).

4. The Rash Rating Model

The Rush model for one-dimensional estimation is an extension of the Rush model of two-dimensional items, which Masters (1982) developed to accommodate multiple staging items. The rush model assumes that all paragraphs are equally distinguished, the prediction is assumed to be zero, and the paragraph's difficulty takes variable values. In the case of measuring a specific ability such as achievement, the difference will be in the item difficulty parameter and in the threshold of difficulty factor for transition from the lower grading to the higher grading within a single item, and any item that has several transition thresholds equal to the number of gradations -1. In the case of items on a Likert scale, it assumes that all the paragraphs have the same difficulty coefficient, but the difference is in the thresholds characteristic of the transition from the lower to the higher level of power, as is the case in the particular model that measures trends or tendencies according to Likert. The mathematical formula for the probability that the individual *(j)* with the ability *(Ø)* responds to the two-stage item *(i)* is a correct answer through the relationship between the individual's ability and the difficulty of the item as follows:

$$P(\emptyset) = \frac{p1ji}{p0ji+p1ji} = \frac{e^{(\emptyset j - bi1)}}{1+e^{(\emptyset j - bi1)}}$$

Where:

(P1ji) indicates the probability that the individual *(j)* will answer the first step in item *(i)* a correct answer.

(P0ji) refers to the possibility that the individual *(j)* will answer the first step in item *(i)* by an incorrect answer.

(Øj) denotes the ability of an individual *(j)*

(bi1) the discriminatory ability of the first step in *(i)*. In this case, there is one step, and therefore: *p0ji + p1ji = 1*.

Masters (1982) argues that the Rush Model for single-category estimation does not necessarily have the threshold difficulty coefficients sequentially within the graded answer model's category boundaries. The difficulty of the first step may be greater than the difficulty of the next step. The partial grade model is an extension of the one-category model; it is assumed that the item does not differ in the discrimination coefficients and that the prediction parameter has a value of zero. When using this form, the test taker needs to have its items structured in a disciplined manner. In other words, the method of answering these items must be in the form of sequential steps, meaning that the subject cannot move to a higher response level without reaching the level that precedes it.

The differential thresholds degree to which the individual possesses the trait, allowing him to choose the higher alternative instead of the lower-level choice.

Methods of building a self-confidence scale: The methods of constructing psychological scales vary according to the general trends for measuring them. There are four directions for measuring self-confidence (Salman, 2007):

First: The method of self-report

Self-reporting measures are based on the direction that looks at the characteristic to be measured as the individual himself perceives it. This is because the individual can express what he feels internally. Therefore, this method is considered one of the best methods for building metrics to measure an individual's internal state (Hammouda & Imam, 1994). It is expected that the individual assesses his characteristic, feelings, or behavior through measures that include several items around the characteristic or behavior to be measured. The scale is designed to enable the individual to express his traits sometimes in writing or orally (Cronbach, 1970).

Self-report scales are more commonly used than others self-rating and performance projective scales. This is due to the ease of preparation and correction, the possibility of codifying them for large samples, and the controllability of many items that can include each feature as much as possible (Al-Kubaisi, 1978). There are three methods used to build the items of the self-report measures: declarative or interrogative expressions, verbal positions, and forced selection. However, the best most common of these methods is the declarative statements method because it is easy to design and prepare and covers the characteristic to be measured to a large extent. This method follows a continuum of five ranking levels for each item following the Likert ranking scale (Odeh, 2010).

Second: The method of respecting others

The evaluation of respecting other scales is a tool that consists of a group of items related to the measured trait. Each item expresses simple behavior that is subject to a gradation of several predetermined levels in proportion to the characteristic to be measured and the age group applied to it (Odeh, 2010). This method is based on building standards that capture the characteristic to be measured as observed in an individual by others. Therefore, others estimate an individual's trait or behavior using graded scales. The scale designer assesses behavioral patterns or

traits through prolonged contact with the subject (Thorndike & Huogen, 1989). This is done by using the order of the stimuli or alternatives or pairwise comparisons between two stimuli or grading the trait. The grading of measures of the evaluation of respecting others is either quantitative or qualitative (Allam, 2009).

Third: the performative or practical method

This type of construction of scales relies on observing the behavior of the individual. The trait to be measured is captured by exposing the individuals to specific and realistic situations prepared in advance for this purpose.

Fourth: the projective method

In this type of scale, the individual reflects his thoughts, trends, or fears on the test material, scale, or tool to be designed. The participant is exposed to some mysterious stimuli that trigger his imagination, revealing the soul's secrets (Faeq & Abdel Qader, 1972).

5. Methods of Building Psychological Measures

There are three basic ways to construct any scale that measures a psychological trait:

First: the logical method: it is a method based on theoretical, logical, or intuitive foundations that define the trait's concept and the various behavioral situations; this means that once we define the character or traits to be measured, we formulate items or paragraphs that measure that designated characteristic or traits (Allam, 2009).

Second: the (experimental) empirical method: the practical measures depend on their construction on their ability to distinguish between people characterized by specific characteristics such as mentally ill persons and others who do not have these characteristics (normal) through experimentation. Third: the factor analysis method: in this method, factor analysis is used to collect the scale paragraphs in the least possible number of factors that explain the variation in the performance of individuals on the scale.

To measure the self-confidence of high school students in Jordan, a measure must be designed and constructed based on a robust theory. As with any psychological newly constructed measure, it needs a certain theory to be built upon. Hence, the psychometric properties of the scale will be verified in light of the theoretical assumptions. For this reason, the researchers in this study adopted the modern theory of measurement (Item Response Theory) to construct the scale of self-confidence. This theory contributed to the development of new psychometric models to ensure the objectivity of the scale, which makes it unaffected by the tool's differences or the participants' individual differences. This theory included several models that allow objective measurement that focuses on the test items. Hence, this allows adding, deleting, or modifying items in the test or scale without affecting the test as a whole. Following is an explanation of the Item Response Theory (IRT).

Item Response Theory (IRT)

This theory came to respond to the problems and shortcomings that faced the classical measurement theory (Allam, 2000). A set of mathematical models emerged from this theory, each of which has its equation that determines the relationship of the individual's performance on the item and the ability that explains this performance (Hambelton & Swaminthan, 1985). These different models allow for objective measurement that focuses on the test item (Kazem, 1996).

These models assume that the measured trait is a specific ability or characteristic of the individual tested with it. Therefore, there is a stable relationship between the measured trait levels in a group of individuals and the probabilities of the correct answer for several different items. Therefore, these models are considered probabilities based on the logarithmic weighting function instead of the equilibrium probability density function on which the classical theory depends (Allam, 2000).

Assumptions of response paragraph theory (IRT)

First: Unidimensionality

Most of the item response theory models assume a single ability or trait that explains the individual's performance on the test or scale. This hypothesis can be achieved when there is one dominant trait. The hypothesis can be verified in two ways: The first depends on determining the dimension to be measured and choosing the items that the dimension corresponds to. The second method is based on determining the scope of the test item's content or the scale and then using factor analysis of the subjects' responses to the items. Then, determining the values of the latent root and the variance ratios explained for the first and second factors, where it can be ascertained of one-dimensional or the presence of one dominant element on the test items when the difference between the value of the first latent root and the values of the latent roots of the rest of the factors is relatively significant (Hambelton & Swaminthan, 1985).

Second: Local Independence

Local independence means that the individual's scores are statistically independent of one another at any point on the trait continuum. This means that the individual's response to any paragraph of the scale is not affected by his response to the other sections (Rash, 1961). Therefore, this assumption demonstrates that the participant's ability and characteristics are the only criteria that affect the individual's performance on this item (Hambelton & Swaminthan, 1985).

Third: Item Characteristic Curve

It is a mathematical function that shows the relationship between the probability of a correct response and the latent ability that the scale measures. Since individuals are different in their latent ability, responding correctly to each item will be different. Consequently, the item characteristic curve takes the form of a logarithmic weighting curve (Bhkata et al., 2005).

The distinctive characteristic of the item characteristic curve

First: item difficulty parameter *(bi)*

It is defined as the power estimate *(Ø)* corresponding to the probability of the correct answer (0.50) when the point of intersection of the item characteristics curve with the y-axis is approximately zero (that is, the parameter of the prediction equals zero). However, suppose the value of the y-section of the item characteristics curve is greater than zero. In that case, the item difficulty is the power represented on the x-axis that corresponds to the probability of a correct answer midway between the intersection of the curve and the y-axis and the value one integer. The difficulty of the item in the modern theory of measurement takes values between (- infinity, + infinity), but in general, the values that it takes are between (-3, + 3) (Harris, 1989).

Second: Item Discrimination *(ai)*

It is an indicator that links a change in individuals' abilities to a change in the correct answer's probability. Discrimination does not mean the slope of the curve but rather a proportion of it; however, as the curve increases, the discrimination increases. Item discrimination expresses the degree to which the item is able to distinguish between the different levels of ability of individuals. This distinction is directly proportional to the slope of the item characteristics curve at the point of inflection. The extent of discrimination in general is (-infinity, + infinity). This means that the possibility of the items with negative discrimination and paragraph items with positive discrimination are always deleted from the ability tests. The expected range of item discrimination parameter falls in the closed period (0, +2) (Hambelton & Swaminthan, 1985).

Third: Item Guessing *(Ci)*

Item guessing expresses the probability of answering an item with a correct answer by the subjects with low ability on very difficult or medium difficulty items. It represents they-segment of the item characteristics curve, and a paragraph can be considered good when the lower asymptote approaches zero; this means that the guess is equal to zero. Although the theoretical value of the guess ranges between zero and the integer one, in practice, its value ranges between zero and 0.25 (Harris, 1989).

Capacity estimation in item response theory:

The ability can be estimated in one of the following two methods: The Maximum Likelihood Estimation method, which depends on the pattern of the subject's response to the test or scale items. If the respondent answered correctly, he scores (1), and if he makes a mistake, he scores (0). The maximum likelihood function is calculated at more than one value for power, and the power value (Ø) that receives the most significant value of the maximum likelihood function is the power of the subject in that test. The greater the number of test items and the greater the number of examiners, the closer the estimated power's value is to the true capacity (Embertson & Reise, 2000). Additionally, the second method is Bayesian Model Estimation which is adopted when it is difficult to use Maximum Likelihood Estimation. This method assumes the existence of a prior distribution of the capacity of the subjects in the community, and in most cases, this distribution is normal with an arithmetic condition (0) and a standard deviation (1) (Warm, 1978).

Estimation of Item Parameters:

When using item response theory models, it is possible to arrive at fixed parameters of the items, which do not depend on the subjects' sample in which the items were evaluated; this means that the item response theory provided a measure free of the sample when estimating the paragraphs' parameters. Thus, the scale is freed from the items when assessing the parameter of the subject's ability. This theory provides us with information that the item must give at a certain level of power depending on the fixed parameters and with a variable standard error of

the change of the subject's ability. There are several methods for estimating these parameters of the item. Multiple ways are used through programs that can be applied using computers (Hambelton et al., 1991).

There are many methods to estimate the items' parameters: the first one is the Maximum Likelihood Joint method, which is one of the most employed methods for estimating items' parameters. There are many easy-to-implement computer programs through which an estimate of the item's parameters is done through two stages. For example, when using the three-parameter logistic model, item parameters are dealt with in the first stage to estimate the subjects' capabilities, and the ability values are used in the second stage to assess the item parameters.

Another method is the conditional maximum likelihood method. This method is valid in the single parameter because it depends on the statistical availability of sufficient information associated with the ability. The statistician is sufficient to estimate the items' parameters in the absence of other information from the data. In the case of the two and three-parameter model, although it is not considered statistically sufficient for the estimation. But it rather depends on the items in which the subject succeeded, and one of the methods for estimating the parameters of the paragraphs is the Marginal Maximum Likelihood method. This method does not depend on the parameters associated with the subject's ability as in the previous method. They are now distributed normally with a mean equal to zero and a standard deviation of one unit. The subjects' data are treated as a random sample from the marginal odds method population (Hambleton & Swaminthan, 1985).

Evaluating level of matching

To test the conformity of the data for the used model, a comparison is carried out by comparing the model's expectations with the observed data. Item mismatch, mismatch of individuals, or both are basic sources for mismatch of the experimental data with the used model's assumptions. If derived data of an item do not match the model expectations, then it does not match the model and should be deleted. This is due to its unstable relative difficulty about the rest of the items across the different levels of the subjects' ability or because it does not belong to the test items. As for individuals, their non-conformity with the model is because the vertebrae's relative difficulty in these individuals is different from these vertebrae's relative difficulty in most of the subjects (Kazem, 1996).

6. Previous Studies

According to the modern theory of measurement (IRT), the researchers of this study did not find (to their knowledge) previous studies on building a scale of self-confidence according to the modern theory of measurement (IRT). Therefore, the previous studies that will be reviewed here can be classified into two parts:

a. Studies dealing with building scales according to the modern theory of measurement (IRT). B. Studies dealing with building scales according to the classical theory of measurement (CTT).

According to the item response theory, Abu Al-Sil (2016) conducted a study to build a scale of motivation of achievement for high school students in Damascus. The scale included 33 items. To ensure the apparent validity, it was presented to a group of specialized arbitrators. The scale was applied to the study sample of (1200) male and female students. The most important results were that the motivation of achievement scale prepared by the researcher was one-dimensional, and this data was identical to the partial estimate model. The average match of individuals and the internal and external items approached zero, and the standard deviation approached one correct.

According to the item response theory, Abu El-Sell (2011) also conducted a study to build a scale of personality types for students at Damascus University. The study sample consisted of 1600 university students. Riso and Hudson's scale was used as a test. The most important results of the study were that the correlation coefficients between the two scales were high. Several experts reviewed the scale to verify its content validity. The researcher used factor analysis to verify the unidimensional personality styles scale. Many individuals matched the model, which was used to estimate the differential thresholds and compare most items through the items and individuals' internal and external indicators of conformity.

Jabara (2007) conducted a study to build a multi-dimensional personality scale according to the modern theory of measurement. The study sample consisted of 450 students from the Faculties of Medicine, Engineering, and Law at several Jordanian universities. This study's findings highlighted eight factors resulting from the factor analysis (balance and rationality, responsibility and motivation for achievement, sadness, decision-making, social capacity, creativity, self-control, and order) as representative of personality dimensions. The researcher verified the validity of the scale by presenting it to eight experts. According to the Cronbach Alpha method's reliability, the half partition method and the coefficients were corrected with the Spearman-Brown equation.

Several studies dealt with building scales according to the classical theory. According to the triple classification,

Abdullah (2003) built a scale to measure the trends of goals of achievement, which targeted university students. The scale consisted of (24) items distributed into three dimensions, consisting of eight items. The response continuum consisted of four alternatives ranging from (fully applicable) to (not fully functional). The scale was of good reliability and validity. The scale achieved validity of the internal consistency and reliability using the Cronbach Alpha method.

Urdan and Midgley (2003) prepared a scale considered one of the most widely used scales in the triple classification framework. The reason for that was that its psychometric properties were good. The scale achieved good validity and reliability. The scale in its final form consisted of (14) items distributed as follows: (5 items that measure the goals of mastery, five items that measure performance goals - courage, and four items that measure performance goals -reluctance). The available alternatives were as follows: it applies completely, it applies, it does not apply moderately, it does not apply, it does not apply entirely.

Study procedures Study Approach:

The researchers used the descriptive analytical approach for its compatibility with the objectives of the study.

Study population and sample:

The study population consisted of high school students in the Al-Mafraq Governorate in the Hashemite Kingdom of Jordan. The population size reached (9760) male and female students, distributed over (74) secondary schools, of which (35) schools were for males and (39) schools were for females. To verify the psychometric properties of the self-confidence scale that was designed, the scale was applied to an experimental sample of (310) male and female high school students in Al-Mafraq Governorate.

The final sample in this study amounted to (1060) male and female students from the secondary stage at Al-Mafraq Governorate schools. The sample was chosen by the random cluster method, so (26) secondary schools were randomly selected from the 74 schools, i.e. (12) male schools and (14) female schools. From each school, two classes were selected randomly from the secondary classes in each school, where the number of males reached (488) students and (572) female students.

Steps of constructing the scale:

The scale was constructed according to the following steps:

First: definition of self-confidence

The theoretical literature related to self-confidence research was reviewed, and many previous studies related to self-confidence were discussed (e.g., Al-Rifai, 2004; Perry, 2011). The researchers suggested the following definition of self-confidence: (it is a trait of personality by which the individuals trust their own judgment and abilities) and value themselves and feel worthy, regardless of any imperfections or what others may believe about them).

Second: Preparing the scale Items and how to answer them

After the self-confidence trait was defined and the basic components were identified, fifty items were formulated. Each item had five alternatives, representing the extent to which the content of each item applies to the respondent: (it applies to me to a considerable time, it applies to me to a large extent, it applies to me to a moderate degree, it applies to me to a slight degree, it does not apply to me at all). So, these items were reviewed by nine experts who were specialists in psychology, statistics, measurement and evaluation, and supervision. In light of the experts' observations, some items were modified or deleted. The agreement among the experts approached (80%). Thus, the number of items on the scale became (44), which formed the scale's initial version.

Third: the instructions and clarity of the scale items

The first version of the scale was administered to a sample of (310) students to determine the students' observations and comments on the scale items in terms of clarity of their content. Besides, this pilot study helped us to determine the appropriate time to apply the scale. This pilot study revealed that all items were clear and understandable, and there were no ambiguous items. Therefore, no item was modified. The pilot study showed that the average time needed to answer all scale items was about (25) minutes.

Scale grading:

The scale was graded as follows: it applies to me to a considerable extent takes five points; it applies to me to a large area takes four moments; it applies to me with an average extent takes three points; it applies to me to a slight degree takes two points; it does not apply to me at all takes one moment. All scale paragraphs are positive, so the scale scores ranged between (44) points as a minimum to (220) points as maximum.

Psychometric properties of the scale

The scale was applied to the pilot sample, which consisted of (310) students. The reliability was calculated using the Cronbach Alpha method, and its value was (0.89), which is a good value. The concurrent validity was calculated by applying the Self-Confidence Scale of Abu Allam (1978) as a test. Hence, the correlation coefficient was calculated between individuals' scores on the two scales, i.e., the scale of this study and the Abu Allam scale. The correlation coefficient value between the two scales was (0.79), which indicated a good correlation between the two scales. Consequently, this confirmed the concurrent validity of our scale. After that, the scale was applied to the final sample of (1060) male and female students to answer this study's research questions.

7. Results

Q1: What are the unidimensional implications of the items of the self-confidence scale? The assumption was verified using the following indicators:

1- Factor analysis

The researchers used factor analysis by the principal components' method, calculating the items' loading factor on the extracted factors and the factors' Eigenvalue. The four elements were extracted to measure self-confidence, and; each underlying root factor is shown in Table (1) below.

Table 1. Eigenvalue of the four factors extracted for the scale

Factor / Scale	Eigenvalue	The ratio of interpreted contrast
First	3.27	23.60
Second	1.52	11.26
Third	1.41	4.24
Fourth	1.12	1.04
The ratio of (first: second)	2.15	
The ratio of (second: third)	1.08	
The ratio of (third: fourth)	1.25	
The ratio of (second: third): (third: fourth)	2.00	

It can be noted from the results of Table (1) that the value of the Eigenvalue of the first factor is (3.27), which explains (23.60%) of the total variance. The test is considered unidimensional if the percentage of what is explained by the first factor is more than (20%) (Reckase, 1979). Therefore, it is evident that the test fulfills the Unidimensionality assumption. The unidimensional factor analysis is adopted through the Eigenvalue ratio of the first factor to the second Eigenvalue. So, the unidimensional factor is achieved when the balance is greater than 2 (Hambleton & Swaminthan, 1985). Therefore, it is clear from the previous Table that the ratio of the first factor's Eigenvalue concerning the Eigenvalue of the second factor is equal to 2.15, which is an indication of the unidimensional aspect of the test items.

2- Internal Consistency Reliability

The high value of the coefficient of internal consistency reliability (KR-20), which amounted to (0.89) for the test model, showed any indication of the fulfillment of the Unidimensionality assumption according to the opinion of Cronbach (1951), who believes that the parameter (KR-20) is a good indicator for verification of Unidimensionality. This is because it represents the expected value of the variance ratio explained by common factors between items when two random samples are linked from the clustering of items. This high value of this factor provides a good indication for verifying the Unidimensionality assumption of the test.

Q2: What are the implications of matching the Self-Confidence Scale's empirical data for the unidimensional partial estimate model?

The compatibility of the individuals to the adopted model was verified to answer this question. Table (2) below shows the index of internal and external congruence for the individuals' spacing's stability to measure self-confidence.

Table 2. Index of internal and external matching and reliability of individuals' dispersion

Scale	matching individuals	INFIT		OUTFIT		Stability of individuals dispersion
		MNSQ	ZSTD	MNSQ	ZSTD	
self-confidence	968	0.18-	1.05	0.14-	1.03	0.84

It is evident from Table (2) that the values of the conformity index for external and internal individuals have an average approaching zero, and their standard deviation is close to the correct one, which indicates that the students (individuals) to whom the scale was applied are identical to the model used in this study. The internal and external congruence of the items and the stability of item dispersion for the self-confidence scale will be explained in Table (3).

Table 3. Index of internal and external matching and reliability of items dispersion

Scale	matching items	INFIT		OUTFIT		Stability of items dispersion
		MNSQ	ZSTD	MNSQ	ZSTD	
self-confidence	39	0.14-	1.01	0.11	1.02	0.94

It is evident from the previous Table that the respective values of the internal and external matching index of the items of the scale have an average of very close to zero, and their standard deviation is close to the correct one. Several (5) items were not identical to the model and were deleted, while the rest of the self-confidence scale items (39) were identical to the model.

Q3: What are the estimated threshold values characteristic of the paragraph scoring levels for the Self-Esteem Scale?

To answer this research question, the researchers used the statistical program WINSTEPS for calculating these concepts to estimate the values of the differential thresholds of the different levels of measurement for the self-confidence sale, as illustrated in Table (4) below.

Table 4. Thresholds values for measurement levels of self-confidence scale items

Scale / Levels	1	2	3	4	5
self-confidence	-2.29	-1.41	-0.53	0.59	1.42

The previous Table regarding the values of the differential thresholds for different self-confidence scale levels shows that these thresholds reflect the differentiation of the participants' responses on the self-confidence scale. Using this scale, the student whose self-confidence is high can be distinguished from the student whose self-confidence is low.

Q4: What are the performance standards for high school students on the scale of self-confidence in terms of transferred logistical units?

To answer this question, the researchers used the WINSTEPS program to estimate individuals' features on the self-confidence scale and the percentile rank corresponding to each value calculated as presented in Table (5) below.

Table 5. The raw scores for each student on the measure of self-confidence and the values of the marks estimated in the logistic unit, and the percentile rank corresponding to each raw score

Raw score	Transferred score	Percentile rank	Raw score	Transferred score	Percentile rank
39	2.88-	0.10	118	1.29	69.50
42	2.60-	0.50	120	1.31	71.50
45	2.22-	1.00	121	1.33	72.00
49	2.10-	1.40	126	1.36	72.50
51	1.98-	2.00	128	1.39	74.50
53	1.94-	2.50	130	1.42	75.00
54	1.82-	3.50	133	1.49	76.00

56	1.65-	5.50	134	1.55	76.50
59	1.55-	6.00	137	1.60	77.00
61	1.44-	7.50	141	1.62	78.00
63	1.39-	9.00	142	1.63	80.00
68	1.12-	11.00	146	1.66	81.50
70	1.11-	12.00	148	1.69	82.00
73	1.01-	14.50	150	1.73	83.00
74	0.88-	18.00	154	1.76	83.50
77	0.71-	23.50	155	1.81	84.00
79	0.62-	29.00	158	1.85	85.00
82	0.55-	35.00	161	1.88	86.00
85	0.43-	41.00	166	1.91	87.00
88	0.33-	46.00	168	1.94	87.50
89	0.21-	50.50	171	1.98	88.00
91	0.15-	52.00	176	2.01	89.00
94	0.09-	53.50	177	2.02	89.50
97	0.02-	55.00	179	2.09	90.00
100	0.17	57.00	180	2.11	91.00
102	0.22	59.00	182	2.23	92.00
103	0.39	60.50	184	2.29	93.00
105	0.96	64.00	186	2.33	95.00
108	0.99	65.00	188	2.39	97.00
110	1.03	66.50	190	2.41	98.00
113	1.09	68.00	192	2.59	99.00
116	1.22	69.00	193	2.77	100.00

It is clear from the previous Table that the raw scores were distributed between 39 to 193, where the lowest score for the scale was (39), and the highest score was 193. As for the converted score, the smallest value was (-2.88), and the highest value was equal to (2.77), which is within the acceptable range according to the item response theory. As for the percentile ranks corresponding to each raw mark, it was within the range (0.10 to 100.00), which indicates that the scale built based on the item response theory measured the self-confidence of high school students in Al-Mafraq Governorate in Jordan.

8. Discussion

The results of this study showed that the items of the self-confidence scale were loaded on the factors extracted from the factor analysis, which was unidimensional, indicating that the scale measures one dimension, which is self-confidence. This was in agreement with the study of Abu Tuberculosis (2016). The results showed indications that a very large proportion of individuals matched the model that was used in the study to estimate the parameters of the differential thresholds, as 93% of the study individuals matched the model used in the study, which is an excellent percentage compared with the number of the study sample (1060) students, as well as matching most of the scale items through the internal and external congruence. Besides, the item's dispersion stability was high, which applies to individuals who matched the model used through the internal and external matching indicators.

This indicates that the scale has good empirical data matching properties. As for the item dispersion's stability, its value was 0.94, which is a very high value. The reason is that the items' dispersion threshold is more constant than the items themselves' dispersion. Therefore, the scale items are identical, and the distribution of the differentiating threshold values that were achieved for the self-confidence scale had a high discriminatory ability to distinguish between the student who has high self-confidence and those who have low self-confidence. Based on these results, the self-confidence scale built in this study can be used to determine the self-confidence of high school students in Al-Mafraq Governorate in Jordan using the partial estimate model in the item response theory (IRT). These findings are in harmony with the findings obtained from the study of Abu Al-Sell (2011).

As for the percentile ranks results, there was a difference in these ranks corresponding to each raw degree, which is a good indication of the effectiveness of the scale items that were constructed. The study results further showed that the logistic-converted marks are within the reasonable range in the item response theory, which enhances our scale's ability to measure the self-confidence of high school students in the Al-Mafraq Governorate in Jordan. Consequently, the possibility of generalizing the results to high school students in Jordan is feasible.

Study limitations and recommendations

The study was limited to high school students in the Al-Mafraq Governorate schools for the academic year 2020-2021. Based on the study's findings, the researchers recommend applying the scale to high school students in other cities or countries to reveal their self-confidence. Thus, the concerned parties could have a general picture of the level of students' self-confidence. This can help decision-makers to provide student sponsorship programs or appropriate mentoring for any low-confidence students. The researchers are also recommended to use the item response theory to design a scale for creativity and innovation among university students and direct graduate students and those interested in psychological and personality traits towards building scales that measure personality traits and characteristics.

Self-Confidence Scale (final version)

Dear Student

Please read each statement carefully and respond as you feel. Remember that there is no correct or false answer, as the correct answer is the answer that expresses the truth about how you think about these points.

First: Personal information

Student Stream: (   ) scientific   /   (   ) literary

Sex: (   ) male   /   (   ) Female

Second: scale items

No.	Items	Apply to a huge extent	Apply to a large extent	Apply to a medium extent	Apply to a low extent	Don't apply at all
1	I hesitate to make the decision even after I'm convinced about it.					
2	I have a set of realistic goals.					
3	I face the results of my behavior.					
4	I have a firm attitude and principle towards what happens to me.					
5	I'm talking to others about my job and my career future.					
6	I love challenging tasks.					
7	I'm constantly trying to improve my abilities and skills.					
8	I apologize to those I have wronged, even if it is in front of others.					
9	I neglect social responsibilities, so I don't deal with others.					
10	I can get through the obstacles I'm facing.					
11	My colleagues cause me a little discomfort repeatedly.					

12	I stop thinking every time I have negative thoughts.
13	I avoid talking to strangers.
14	I take the risk for an idea he's convinced of.
15	I feel that others accept my appearance.
16	I keep fit through daily exercises.
17	I evaluate my thoughts and actions objectively.
18	I accept constructive criticism from my colleagues.
19	I have a determination to overcome difficulties.
20	I switch the things I buy many times.
21	I ask for help from others in my life so much.
22	I make new friends quickly.
23	I accept advice from others, even if it goes against my thoughts.
24	I feel like my ability to influence others is weak.
25	I'm afraid to put my thoughts to others.
26	I'm trying to get used to other people's experiences.
27	I have social attitudes without hesitation.
28	I change my mind quickly when others oppose it.
29	I'm afraid to admit my mistakes.
30	I help my colleagues solve their problems.
31	I can achieve my goals, even though it's hard.
32	I expect success in most of the work I do.
33	I can put my thoughts in front of others.
34	I plan my future with confidence.
35	I persevere to achieve my goals.
36	I win the trust of others quickly.
37	I treat others with caution so I don't get criticized.
38	I hesitate to inquire about the mysterious issues.
39	I think I'm a particular person compared to others.

References

Abdullah, R. (2003). Differences between students with high and low achievement in self-organized learning strategies and goal orientations in a sample of students from the Faculty of Education in Oman. *Journal of Psychological and Educational Research, 2*, 13-98.

Abu Allam, A. (1978). *Measuring self-confidence among female students in secondary and university levels*. Kuwait: Al-Sabah for printing and publishing.

Abu Al-Sell, M. (2011). *According to the IRT's partial assessment models, building a personality patterns scale and verifying its effectiveness for university students*. Unpublished Ph.D., Arab University of Amman, Amman, Jordan.

Abu Al-Sell, M. (2016). Building a scale of achievement motivation for high school students in Damascus in accordance with the IRT. *Journal of the Union of Arab Universities for Education and Psychology, 4*(4), 140-175.

Al-Anzi, F. (2001). The sub-components of self-confidence and shyness are a practical correlational study. *Journal of Social Sciences, 3*(29).

Al-Badrani, J. (1986). *Building a self-confidence scale among middle school students*. Unpublished master's thesis. The University of Baghdad. Iraq.

Ali, S. (2009). *The effectiveness of a proposed mentoring program for developing self-confidence among students of the Islamic University of Gaza*. Gaza. Islamic University.

Al-Kubaisi, K. (1987). *Building and codifying a scale of priority personality of admission to military colleges for sixth graders in Iraq*. Unpublished Doctoral thesis, University of Baghdad. Iraq.

Allam, S. (1986). *Contemporary developments in psychological and educational measurement*. Kuwait: Commercial presses.

Allam, S. (2000). *Analysis of psychological, educational and social research data, i3*. Cairo: Arab Thought House.

Allam, S. (2009). *Measurement and evaluation in the teaching process*. Dar al-Marcha: Amman: Jordan.

Al-Omar, B. (2000). The relationship of motivation to work with some personal and functional variables among employees in the state of Kuwait. *Journal of the Center for Educational Research, 5*(17), 56-77.

Al-Rifai. (2004). *The impact of a proposed psychological mentoring program on the development of self-confidence and its relationship to the level of athletic achievement*. Iraq.

Al-Rikabi, N. (2000). *The level of ambition and its relationship to self-confidence among the students of the Faculty of Education*. An unpublished master's thesis. Mustansiriya University. Iraq.

Barron, F. (1953). An ego strength scale to predict response to psychotherapy. *Journal of Consulting Psychology, 17*, 327-333. https://doi.org/10.1037/h0061962

Bhakta, B., Tennent, A., Horton, M., & Lawton, G. (2005). Using Item Response Theory to Explore the Psychometric Properties of Extended Matching Question Examination in Undergraduate Medical Education. *BMC Medical Education, 5*(9). https://doi.org/10.1186/1472-6920-5-9

Bunker, L. K. (1991). The role of play and motor skill development in building children's self-confidence and self-esteem. *The Elementary School Journal, 91*(5), 476-471. https://doi.org/10.1086/461669

Cottle, T. J. (1974). The Felt Sense of Studentry. *Interchange, 5*(2), 1-41. https://doi.org/10.1007/BF02138827

Cronbach, L. (1971). Test Validation. In R. L. Thorndike (Ed.), *Educational Measurement* (pp. 443-507). Washington, DC: American Council on Education.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334. https://doi.org/10.1007/BF02310555

Cronbach, L. G. (1970). *Essentials of Psychological Testing*. New York, Harper Row.

Faiq, A., & Abdul-Qader, M. (1972). *Introduction to General Psychology*. Cairo: Anglo-Egyptian Library.

Fennema, E., & Sherman, J. A. (1978). Fennema-Sherman Mathematics Attitudes Scales: Instruments designed to measure attitudes toward mathematics learning by males and females. *Journal for Research in Mathematics Education, 7*, 324-326. https://doi.org/10.5951/jresematheduc.7.5.0324

Geist. C., & Hamrick. T. (1993). Social avoidance and distress: its relationship to self-confidence, and needs to affiliation, change dominance, and deference. *Journal of clinical psychology, 39*(5), 727-730. https://doi.org/10.1002/1097-4679(198309)39:5<727::AID-JCLP2270390514>3.0.CO;2-F

Hambelton, R., & Swaminthan, H. (1985). *Item response theory: Principles & Applications*. Norwell, MA: Kluwer Nijhoff Publishing. https://doi.org/10.1007/978-94-017-1988-9

Hambleton, R. K., Swanminathan, H., & Rogers, H. (1991). *Fundamentals of Item Response Theory*. New York: Sage Publications the International Professional Publishers.

Hamouda, A. R., & Imam, A. H. (1994). Measures of adolescents and adults' psychological state. *Journal of Psychology: Cairo, 3*(31), 6-35.

Harris, D. (1989). Comparison of 1-, 2- and 3- Parameter I R T Models. *Educational Measurement: Issues and Practice, 8*(1), 35-41. Maryland: Universities of Maryland, U.S.A. https://doi.org/10.1111/j.1745-3992.1989.tb00313.x

Jabbara, K. (2007). *Building a multi-dimensional scale and verifying its effectiveness in predicting academic achievement in the faculties of medicine, engineering, and law*. An unpublished Ph.D. thesis, Arab University of Amman, Amman, Jordan.

Kazem, Mohammed. (1996). *Use the Rush model to build an initial scale in psychology and achieve objective interpretation of the results*. Cairo: Anglo-Egyptian Library.

Kleinmuntz, B. (1960). Identification of Maladjusted College Students. *Journal of Counseling Psychology, 7*(3), 209-211. https://doi.org/10.1037/h0047989

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 2*, 149-174. https://doi.org/10.1007/BF02296272

Nassif, S. (1993). *Social growth and self-confidence among children deprived of parents*. Unpublished master's thesis, Faculty of Education, King Saud University, Riyadh.

Odeh, A. (2010). *Measurement and evaluation in the teaching process, I3*. Jordan: Al Amal Publishing and Distribution House.

Perry, P. (2011). Concept Analysis: Confidence/Self-confidence. *Nursing Forum, 46*(4), 218-230. Blackwell Publishing Inc. https://doi.org/10.1111/j.1744-6198.2011.00230.x

Rash, G. (1961). On General Laws and the Meaning of Measurement Psychology. *Proceeding of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 4*(4), 321-333.

Reckase, D. (1979). Unifactorial Latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4*, 207-230. https://doi.org/10.3102/10769986004003207

Salman, A. S. (2007). *The trade-off between the patterns of the formulation of verbal positions in the light of the psychometric characteristics of psychological measures*. An unpublished Ph.D. thesis, University of Baghdad, Faculty of Education. Iraq.

Thorndyke, R., & Hegen, E. (1989). *Measurement and Evaluation in Psychology and Education* (translated by Abdullah Zaid al-Kilani and Abdul Rahman Adass). Amman: Jordan Book Center.

Urdan, T., & Midgley, C. (2003). Changes in the Perceived Classroom Goal Structure and Patterns of Adaptive Learning During Early Adolescence. *Contemporary Educational Psychology, 28*(4), 524-551. https://doi.org/10.1016/S0361-476X(02)00060-7