InterStat, 2007 April, 8
http://interstat.statjournals.net/YEAR/2007/abstracts/0704008.php

REVIEW OF RELIABILITY AND FACTORS AFFECTING THE RELIABILITY

Ilker ERCAN[1], Berna YAZICI[2], Gokhan OCAKOGLU[1], Deniz SIGIRLI[1], Ismet KAN[1]

[1] Uludag University, Medical Faculty, Department of Biostatistics, Bursa, TURKEY.

[2] Anadolu University, Science Faculty, Department of Statistics, Eskisehir, TURKEY.

Corresponding author: Dr. Ilker ERCAN

Adress: Department of Biostatistics, Medical Faculty, University of Uludag, Gorukle, Bursa, TURKEY

Phone: 90-224-4428200 (21028)

Cellular Phone: 90-505-3781903

Fax number: 90-224-4428666

E-mail adress: ercan@uludag.edu.tr

REVIEW OF RELIABILITY AND FACTORS AFFECTING THE RELIABILITY

Abstract

A reliable and valid measure is required for measuring the abstract characteristics. The reliability of the measures are affected by the length of the scale, definition of the items, homogeneity of the groups, duration of the scale, objectivity in scoring, the conditions of measuring, the explanation of the scale, the characteristics of the items in scale, difficulty of scale, and reliability estimation method. Those factors must be known and taken into account to construct a reliable scale.

Key words: Reliability, measurement, scale

1. Introduction

A scale is needed to measure and that scale must be reliable and valid. It doesn't matter the scale's being reliable in case of measuring the concrete characteristics. But it is an important problem in case of measuring the abstract characteristics. So the reliability of the scale gains importance. As a result of this, the concepts about measuring and reliability and also the factors which affect the reliability must be known while constructing a scale. If the researcher who constructs a scale knows those factors, the reliability of the scale will increase and estimates with minimum variances for the population parameter will be obtained.

2. Scientific Importance of Measuring

If the characteristic that the researcher studying on can be measured and mentioned numerically, those can be evaluated scientifically and can be explained. The characteristics that can not be measured or mentioned numerically can only be described (Armagan 1983; Gumus 1977).

That importance of measuring is one of the general characteristics of the science. Science is an enterprise of mentioning the results precisely, clearly and correctly depending on some assumptions. It is necessary to determine the relationships between the scientific data, to use the data consistently, to take the advantage of mathematical expression and inferences. This depends on that data can be expressed mathematically or the observed data's being measurable. If the observed data can not be mentioned numerically, imprecise results or suspiciously correct results are obtained

with those data. That causes a vicious circle and prevents a development in science (Armagan 1983).

A quick development in science depends on measuring the relative characteristics and mentioning them numerically. The basic problem here is to convert the qualitative concepts to quantitative concepts. This problem mostly can be solved in natural sciences but occurs as a problem in society sciences and even today it is an important subject of discussions. Although those handicaps and difficulties, there are positive historical developments and all branches of science tends numerical expression. Even there are so big developments about this subject that, positivism of sciences are measured with the degree of being numeric of the concepts (Armagan 1983).

3. Concept of Measure

Since the statistical techniques aim to analyze and interpret the data, the first way of getting information about the variables is measuring (Comlekci 1989; Oncu 1994).

Measuring is paring the degree of having a characteristic of the statistic units with a symbol or especially with a number, adapting to some rules.

Scientific researches can only be made by data. Researchers must pay attention to use the numbers that are defined by standard measuring techniques for the data. The measurement tool's being standard measurement tool which will be used in measuring the variable's dimension will provide the data have statistical characteristics (Ozdamar 2002).

It is not the same designating a number or a symbol to the statistic unit that has concrete characteristics, with designating a number or a symbol to the statistic unit

that has abstract characteristics. So in social sciences, it is more difficult to measure than basic sciences or medical sciences (Armagan 1983; Carmines & Zeller 1982; Gursakal 2001).

4. Scales

The measurement class which has clear mathematical characteristics is called scale (Ozdamar 2002). Measuring is a determination process and scale is a tool used for this aim (Yildirim 1983). There are important relationships between scale and statistical analyzing. If a value designated to a variable is defined using objective criteria, the results of statistical analysis of those measurements are highly consistent. Using a standard measurement tool or not; gives the division of the variables' numerical definitions according to objective or subjective criteria (Ozdamar 2002).

Measurement process is not only to give numbers to the statistical units, but also is to give mathematical properties together with those numbers. So while giving the numerical values in measuring process, the mathematical properties must also be taken into account. To know according to which rule the variable is measured, is also necessary while choosing the appropriate technique that will be used in data analysis (Comlekci 1989).

5. The Error Rate in Measuring the Concrete and Abstract Characteristics

In measuring the concrete characteristics, the characteristic that will be measured is invariant or it becomes invariant under some circumstances with the control of a person. Secondly, the scale is stable in itself and what it measures is also

very clear. Thirdly, the rules for comparing the characteristic that will be measured and the measurement tool are easy and those rules can be adapted easily. Providing those conditions make it easier to determine the measurement error rate (Ozcelik 1981).

In measuring the abstract characteristics, measurement process has empirical and theoretical aspects. Empirical aspect occurs in questionnaire, observation study and the outputs of the responses. Theoretical aspect can not be measured directly. This can be estimated from the responses thinking the concepts about the subject. So there is a strong relationship between empirical and theoretical aspects. When this relationship is strong the analysis of empirical aspect will provide getting useful knowledge about theoretical aspect on the related subject (Carmines & Zeller 1982).

The facilities in measuring the concrete characteristics are mostly disappear while measuring the abstract characteristics. These can be listed as follows: (Ozcelik 1981)

i. The abstract characteristics are usually not stable as much as required.

ii. There aren't measurement tools stable enough and also there aren't measurement tools those show clearly in measuring most of the abstract characteristics what is measured.

iii. It is quite complicated to compare the scales that would be used for measuring the abstract characteristics.

iv. It is almost impossible to repeat the abstract measurements on the same units under the same circumstances.

v.The measurement tools for abstract characteristics are usually ordinal or interval scales. So the starting point is arbitrary and they don't demonstrate the magnitude of relative characteristic clear enough.

6. The Properties That a Scale Must Have

The aim of measuring is to make an evaluation about the units, events or objects that are subject to measure and to make decisions depending on the results of evaluation. The correctness and appropriateness of the decision depend on the evaluation results, the measurement results and appropriateness of measurement tool. So it is necessary to use a standard measurement tool.

Until standardizing the measurement quality, the items are analyzed and revised again. Managing, standardizing, scoring and interpreting of the standardized scale must be clearly defined. The scales standardized in this way are called objective scales (Gay 1985).

A scale must be reliable and valid to be a standardized scale and then to be good enough to get appropriate knowledge.

7. Reliability

Reliability, a property that a scale must have, is an indicator of consistency of measurement values obtained from the measurements repeated under the same circumstances (Oncu 1994; Carmines & Zeller 1982; Gay 1985; Sencer & Sencer 1978; Arkin & Colton 1970; O'Connor 1993; Carey 1988; Gursakal 2001). There are two approaches in interpretation of the measurement consistency. First approach is fixed row of the unit in the group in successive measurements, the second approach is the magnitude of the measurement errors in successive measurements, and in other words, obtaining the almost same measurement values from the same subject in successive

measurements, so it is related with getting a small measurement standard error (Thorndike et.al. 1991; Tekin 1977).

Reliability is not only a property of a scale; it is also a property of measurement tool and a property of the results of the tool (Oncu 1994).

In the repeated measurements of the concrete characteristics, much closed results to each other are obtained. In measuring the abstract characteristics, it is very difficult to get the same results. So the scales used for the concrete characteristics are more reliable (Gumus 1977).

There are some differences between the scales used for measuring the concrete characteristics and the others. That kind of scales does not give the results with the magnitude meanings since they have arbitrary starting points even if they have the standard intervals. So, it is necessary to examine the power of giving the closed measurements of the characteristic that the scale is measuring with the real values and in this way it is necessary to be analytical and separate the variation, obtained from the measurement tools, using the following ways: (Ozcelik 1981)

i. the variation sourcing from the real differences in the relative characteristic,

ii. the variation sourcing from the real differences that comes out as a result of the effect of the characteristic that is not required to measure on the characteristic that is required to measure,

iii. the variation sourcing from the interaction of i and ii,

iv. the variation sourcing from the measurement error.

A reliable scale is a scale that makes measurements without an error. Since it is impossible to make a measurement without an error, it is possible to increase the reliability of the measurement by decreasing the error to the least value. The basic

condition of decreasing the error to the least value is to determine the error sources and try to control them (Oncu 1994; Gay 1985; O'Connor 1993). The source of the variation might be the measurement tool or a source other than the measurement tool. It is necessary to research the reliability of the measurement tool using some methods to decrease the error sources about the measurement tool (Dawson & Trapp 2001).

An increase in addition observed measurement error will cause a decrease in variation sourcing from the real differences relative to the characteristic that is measured and that will cause a decrease in reliability (Ozcelik 1981). A decrease in addition observed measurement error will provide the demonstration of the real differences between the units and that will cause an increase in reliability (Oncu 1994; Kucukahmet 2000).

Taking account the real value with two elements (t_i) and measurement error (e_i) observed measurement value (x_i) can be defined as follows taking (Gay 1985; Gursakal 2001; Traub 1994).

$$x_i = t_i + e_i$$

In the application, only the measurement value with the observation value (x_i) can be known. Real value (t_i) and error value (e_i) can not be known. Since the error has a random distribution, it distributes randomly in positive and negative way and so the mean of the measurement error is equal to zero ($E(e_i)=0$). Using that assumption the variance of the measurement can be formulated as follows: (Gay 1985; Gursakal 2001).

$$\sigma_x^2 = \sigma_t^2 + \sigma_e^2$$

σ_x^2 : Variance of the observed values

σ_t^2 : Variance of the real values

σ_e^2 : Variance of the measurement errors

Reliability degree of a measurement tool can be obtained by interpreting the reliability coefficient obtained by dividing the variance of the real values to the variance of the observed values (Gursakal 2001; Traub 1994; Turgut 1993).

$$\rho_x = \frac{\sigma_t^2}{\sigma_x^2}$$

Reliability coefficient is not the measure of the error. It is the measure of the faultless and is not an individual statistic; it is a group statistic (Oncu 1994).

The standard error of measurement, occurring because of the sources other than the relative characteristics, can be obtained using last two equations (Oncu 1994; O'Connor 1993; Traub 1994; Turgut 1993).

$\sigma_t^2 = \sigma_x^2 \rho_x$ $\quad$ σ_e : Standard error of the measurement

$\sigma_e^2 = \sigma_x^2 - \sigma_t^2$ $\quad$ σ_x : Standard deviation of the observed values

$\sigma_e^2 = \sigma_x^2 - \sigma_x^2 \rho_x$ $\quad$ σ_t : Standard deviation of the real values

$\sigma_e^2 = \sigma_x^2 (1 - \rho_x)$ $\quad$ ρ_x : Reliability coefficient

$$\sigma_e = \sigma_x \sqrt{1 - \rho_x}$$

There will be an increase in measurement standard error as the standard deviation of the observed values increase or reliability decreases.

$$\sigma_t^2 / \sigma_x^2$$

It is impossible to know the real ratio above but it is used in theory to calculate the reliability ratio. Methods are developed to calculate the reliability coefficient for different situations (Gay 1985; Gursakal 2001).

The reliability of the scale can be measured by different ways. The reliability of the scale can be measured by using a scale once, by using a scale twice or by using two equal scales once. The reliability coefficient ranges between 0 and 1.

8. The Factors That Affect the Reliability

Reliability is not a property of only the measurement tool; it is also the property of the results of the measurement tool (Oncu 1994). Reliability is the measurement of the faultless. There are some factors affecting the reliability of the result taking from the scale (Tekin 1977). Some of the factors are related with the scale, some others are related with the group the scale is applied, and some others are related with the environment (Gay 1985). So those factors must be taken into account at the stage of constructing a scale and at the stage of application. Some factors affecting the reliability are as follows:

8.1 The Length of the Scale

The length of the scale affects the real values and the variances of the observed values. The measurement errors are smaller in the measurement values obtained from the long scales than the short scales (O'Connor 1993). Because huge number of items present the abstract characteristic better (Gay 1985). In this case the number of the items must be increased to increase the reliability (Oncu 1994; Gay

1985; Thorndike et.al. 1991; Traub 1994). But if the scale is not reliable, to increase the number of the items does not make the scale reliable (Tekin 1977). Spearman-Brown equation is used to calculate how many times increment or decrement in the item number causes how much increment or decrement in the reliability coefficient (Oncu 1994; Gay 1985; O'Connor 1993; Thorndike et.al. 1991; Tekin 1977; Traub 1994).

$$N = \frac{\rho' (1 - \rho)}{\rho (1 - \rho')}$$

ρ : Currently level of reliability

ρ' : Required level of reliability

After a limit, any increase in the item number does not provide any advantage in reliability. Huge number of items may cause tiredness, weariness, and carelessness (Tekin 1977). So the advantage in the reliability and the duration of answering the questions must be taken into account together in increment the number of the items.

8.2. The Expression of the Items in the Scale

To get the reliable information in relative subject depends on the expression of the item as required. If the item isn't expressed as required, it is difficult to get required answers (Sencer & Sencer 1978). The reliability of the scale negatively affected in this case since the answer in different times would be different (Tekin 1977).

The basic problem of getting information is not related with the respondents. The units required for the information do not avoid to answer, to state their situations or opinions in case of expressing the item as required. The observations about the units' attitudes in this subject show that it is a good idea to expect the valid answers except for

the private conditions. In that case, the important point is to prepare the item as required (Sencer & Sencer 1978).

In the application of a scale, items must be prepared due to some rules to provide item-answer relations. While items are being directed one by one, sometimes deviations affecting the answers can be met. Those conditions prevent the expression of the items as required; can be seen basically because of insufficiency, misunderstanding, and biasness (Sencer & Sencer 1978).

8.2.1. Insufficiency

The item is called insufficient if it is prepared that it lets overlooks of the details. The in sufficiency basically may become because of deficiency, having several meanings, and indefiniteness. A defected item is an item deprived of the knowledge it has to. An item with several meanings is an item having more than one meaning and without a limitation in its subject. Indefinite item is an item deprived of the measurement that let to a certain measurement (Sencer & Sencer 1978).

8.2.2. Misunderstanding

Misunderstanding basically sources from the linguistic properties the item has. There are several reasons that may cause misunderstanding of the item: (Sencer & Sencer 1978).

i) The words that are out of knowledge and experiences of the respondents should not be used.

ii) The words in the item should not include various meanings.

iii) Item could be misunderstood in case of including a linguistic problem or in case of not arranged appropriately.

8.2.3. Biasness

Some items have a tendency to get some answers in one way as a result of their way of constructing. Those items, not giving the same chance to all answers, are called biased items. The item types causing biasness can be listed as follows.

a. Directing Items

The items, affecting the respondents and directing the answers to one way, construct that class.

b. Loaded Items

The items with some feeling or meaning in a defined subject and tending an approval and remembrance by itself or the items with sayings.

c. The items that cause prestige biasness

Those items cause respondents seem more superior than they are.

d. The items that cause restlessness

The items that cause restlessness since they research the illegal attitudes and behaviors private subjects or since they require the answers with low prestige construct that class.

8.3. Homogeneity of the Group

Another factor affecting the reliability of the item is homogeneity of the group. When other conditions are equal, the more homogeny group provides the more

reliable scales. That can be seen with a transformation to the equation of the reliability coefficient.

$$\rho_X = \sigma_T^2 / \sigma_X^2 ;$$

$$\rho_X = \frac{\sigma_T^2}{\sigma_X^2}$$

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2$$

$$\rho_X = \frac{\sigma_X^2 - \sigma_E^2}{\sigma_X^2}$$

$$\rho_X = 1 - \frac{\sigma_E^2}{\sigma_X^2}$$

There is no reason for a change in observed values of the subjects depending on the group properties. When σ_E^2 is theoretically taken as square of deviation of the observed and the real values, σ_E^2 must be affected form the group heterogeneities. There will be increment in σ_X^2 depending on the heterogeneity of the group. If σ_E^2 is invariant and σ_X^2 increases ρ_X also increases. So it is important to know the heterogeneity of the group that the reliability is estimated (Oncu 1994).

The magnitude of the reliability coefficient depends on the variance of the real values. As can be seen from the formula $\rho_X = \sigma_T^2 / \sigma_X^2$, if the real value of the measured characteristic does not change among the subjects, reliability will be zero whatever the variance of the observed values is. On the other hand if the subjects are

chosen in the way that may cause an increment in real values' variance, the reliability of the scale tends to increase (Traub 1994).

8.4. The Duration of the Scale

If the scale is prepared to measure in a limited time, the insufficiency of time decreases the reliability of the scale. The time must be enough for respondents to answer all the items (Oncu 1994). Since the limit in time causes excitement and carelessness, the reliability of the scale decreases. In case of insufficiency time in scales, careless answers will be given and that will cause to get values closed to zero for the scale's reliability (Oncu 1994; Carey 1988; Traub 1994; Turgut 1993).

8.5. Objectivity in Scoring

The reliability of a scale is affected from the scoring's being whether objective or the researcher's being whether subjective. The consistency of the scores observed from the same or different subject in different times is called that scale's scoring reliability. If a score obtained from a scale is not changing due to the person scoring the scale or the time of scoring that means the scale's scoring reliability is high. If a scale's scoring reliability is high, scale's reliability will be high tool (Oncu 1994; Tekin 1977).

The scoring reliability depends on the scoring's being objective (Tekin 1977). The items with two ore more than two choices are the items with the highest scoring reliability (Gay 1985; Tekin 1977; Traub 1994).

So the scoring process is important for all scales. Scales must be put in form as objective as possible (Gay 1985).

8.6. The Conditions in Making a Measurement

Respondent's being tired, careless, and sleepless, the atmosphere of the measurement and the temperature will cause unwillingness. This will affect the scale's reliability in negative way (Oncu 1994).

8.7. The Explanation of the Scale

The same expressions must be used in the first page of the scale explanation part so all the respondents will understand the same things. The aim of the scale must be told the respondents clearly; the information about how the scale will be responded and determined the principals will be taken into account about anonymous (Serper & Gursakal 1989).

8.8. The Characteristics of the Items of the Scale

The reliability of the values obtained from a scale must be dependent on the item's characteristic properties. In the items of the scale two characteristic properties are taken into account about the reliability of the measurement values. These properties are "differentiating index" and "reliability index". Other than those two properties, in the scales that aim to measure some other information, another property about the reliability of the measurement values, the difficulty index is taken into account (Traub 1994).

8.8.1. Differentiating index

The correlation between the observed values of the items and the values obtained from the whole scale is called differentiating index (ρ_{XY_i}). That index gives information about the level of being in the same order of the item values and the scale values. From this point of view, if the differentiation index of the relative items is high, that means the item is harmonious with the whole and has a positive addition to the reliability coefficient (Traub 1994).

If the correlation of the values obtained from the whole and the values item taken is low, that means that item is unnecessary in the measurement tool and must be removed from the scale (Ozdamar 2002).

8.8.2. Reliability index

Let's remember the Cronbach's alfa coefficient to examine the reliability index.

$$\alpha = \frac{n}{(n-1)}\left[1 - \frac{\sum_{i=1}^{n}\sigma_{Y_i}^2}{\sigma_x^2}\right]$$

n: number of item

Alfa coefficient is the function of the variances of the item values in the scale ($\sigma_{Y_i}^2$) and the variance of the values obtained from the total of the scale (σ_X^2).

σ_X^2 Transformed in the way denoted below;

$\sigma_X^2 = \sigma_{XX}$

$$= \sigma_X \left[\sum_{i=1}^{n} Y_i \right]$$

$$\rho_{XY_i} = \frac{\sigma_{XY_i}}{\sigma_X \sigma_{Y_i}}$$

$$\sigma_X^2 = \sum_{i=1}^{n} \sigma_X \sigma_{Y_i} \rho_{XY_i}$$

$$\sigma_X = \sum_{i=1}^{n} \sigma_{Y_i} \rho_{XY_i}$$

Obtained σ_X is a result of the item standard deviation (σ_{Y_i}) and item-total correlation coefficient (ρ_{XY_i}: differentiating index) and called as the reliability index of the item ($\sigma_{Y_i} \rho_{XY_i}$). For the each item $\sigma_{Y_i} \rho_{XY_i}$ reliability coefficient and for the whole items $\sum_{i=1}^{n} \sigma_{Y_i} \rho_{XY_i}$ reliability coefficient is examined (Traub 1994).

8.9. Difficulty of the Scale

Another factor is the difficulty degree of the items in the scales that aim the level of the knowledge on subject. The knowledge scales must be prepared appropriately to the knowledge level of respondents. The reliability of the scales with very difficult or very easy items will be low since the variability among the values will be low (Oncu 1994; Traub 1994).

If the items in the scale are very easy, since the items will be answered by all the subjects, the mean of the values obtained from the scale will relatively be high and

the difficulty index will be close to 1. In this case the distribution of the scale scores will be negatively skewed (Oncu 1994; Traub 1994).

If the items in the scale are very difficult, since the items will be answered carelessly by the subjects, the mean of the values obtained from the scale will relatively be low and the difficulty index will be close to 0. In this case the distribution of the scale scores will be positively skewed (Oncu 1994; Traub 1994).

The items must be prepared in a way that, the total scores must not have a skewed distribution, has the possible biggest variance (π_i (1-π_i)) and so maximize the reliability and the difficulty index must be 0.50 (Oncu 1994; Traub 1994).

8.10 The Differences Sourcing from the Reliability Estimation Method

It is natural to have different reliability coefficients obtained using different methods, because the reliability definitions are different depending on the calculation of the reliability coefficient. So reliability is affected from the different sources of variation (Gay 1985; Sencer & Sencer 1978; Thorndike et.al. 1991; Tekin 1977).

The reliability coefficient obtained from the parallel forms applied in different times is also affected from the sources of variation. The variation sourcing from the sampling of the items do not affect the reliability coefficient in the case of measuring in the same time or by intervals. If the scales are applied in one time, the reliability coefficient won't be affected from the variation in person. If the scale is constructed from only one document and it is applied in one time, the reliability coefficient won't be affected from the application speed (Thorndike et.al. 1991; Tekin 1977).

9. Conclusion

The reliable scales must be used to make estimation with minimum variances. Reliability gains more importance in the measurements of the abstract characteristics and in the interpretation of those measurements for the researchers. So the researcher must know about the measuring, reliability and the factors affecting the reliability.

References

Armagan, I. (1983). Yontembilim-2 bilimsel arastirma yontemleri (Methodology-2, methods for scientific research). Izmir: Dokuz Eylul Universitesi Guzel Sanatlar Fak. Yayinlari.

Gumus, B. (1977). Egitimde olcme ve degerlendirme (Measurement and evaluation in education). Ankara: Kalite Matbaasi.

Comlekci, N. (1989). Temel istatistik ilke ve teknikleri (Basic statistics' principles and technics). Eskisehir: Bilim Teknik Yayinevi.

Oncu, H. (1994). Egitimde olcme ve degerlendirme (Measurement and evaluation in education). Ankara: Matser Basim.

Ozdamar, K. (2002). Paket programlarla istatistiksel veri analizi–1 (Statistical data analysis by custom softwares-1). Eskisehir: Kaan Kitabevi.

Carmines, E. G., & Zeller R. A. (1982). Reliability and validity assessment. Beverly Hills: Sage Publications.

Gursakal, N. (2001). Bilgisayar uygulamali istatistik-I (Statistics with computer applications-I). Bursa: Alfa Yayinlari.

Yildirim, C. (1983). Eğitimde olcme ve degerlendirme ogretmenler icin el kitabi, (Measurement and evaluation in education hand book for teachers). Ankara: OSYM Eğitim Yayinlari, Yayin No:7.

Ozcelik, D. A. (1981). Okullarda olcme ve degerlendirme (Measurement and evaluation in the schools). Ankara: USYM-Egitim Yayinlari.

Gay, L. R. (1985). Educational evaluation and measurement. London: A Bell & Howell.

Sencer, M., & Sencer, Y. (1978). Toplumsal arastirmalarda yontembilim (Methodology in community research). Ankara: Dogan Basimevi.

Arkin, H., & Colton, P. R. (1970). Statistical methods. New York: Barnes & Noble Books.

O'Connor, R. (1993). Issues in the measurement of health-related quality of life, working paper 30. Melbourne: NHMRC National Centre for Health Program Evaluation, ISBN: 1-875677-26-7.

Carey, L. M. (1988). Measuring and evaluating school learning. London: Allyn and Bacon.

Gursakal, N., 2001. Sosyal bilimlerde arastirma yontemleri (Research methods in social sciences). Bursa: Uludag Uni. Guclendirme Vakfi Yayini No:178.

Thorndike, R. M., Cunningham, G. K., Thorndike, R. L., & Hagen, E. P. (1991). Measurement and evaluation in psychology and education, fifth edition. New York: Macmillian Publishing.

Tekin, H. (1977). Eğitimde olcme ve degerlendirme (Quantitation and evaluation in education). Ankara: Mars Matbaasi.

Dawson, B., & Trapp, R. G. (2001). Basic & clinical biostatistics, third edition. New York: Lange Medical Books/McGraw International Editions.

Kucukahmet, L. (2000). Ogretimde planlama ve degerlendirme (Planning and evaluating in education). Ankara: XI. Baski, Nobel Yayin Dagitim.

Traub, R. E. (1994). Reliability for the social sciences. London: Sage Publications.

Turgut, F. M. (1993). Egitimde olcme ve degerlendirme metodlari, (Measurement and evaluation methods in education), dokuzuncu baski. Ankara: Saydam Matbaacilik.

Serper, O., & Gursakal N. (1989). Arastirma yontemleri (Research methods). İstanbul: Filiz Kitabevi.