



## Cronbach's Coefficient Alpha: A Meta-Analysis Study\*

### Cronbach Alfa Katsayısı: Bir Meta Analizi Çalışması

Hatice İNAL\*\*, Esin YILMAZ KOĞAR\*\*\*, Eda DEMİRDÜZEN\*\*\*\*, Selahattin GELBAL\*\*\*\*\*

• Received: 10.03.2016 • Accepted: 28.03.2016 • Published: 31.01.2017

**ABSTRACT:** This study is to examine the meta-analysis results acquired from Cronbach alpha reliability coefficient being used in Hacettepe University Journal of Education. Within this context, 1222 items taking place in 43 issues which were published in Hacettepe University Journal of Education between 1986 and 2012 were examined and 354 measurement tools in total were discussed according to their inclusion criteria. In this study, r index was used in calculating influence quantity in correlational data for combining data while random influence model of Fisher z method was used in correlational data for combining influence quantities. It was benefited from SPSS 20 and MetaWin 2.0 packet programs for the analysis of data. It was found in this study that effect size mean is quite strong according to various moderator varieties of Cronbach alpha reliability coefficient. As a result of this study, it was determined that effect size mean of scale reliability prepared for measuring affective structure was greater than the scales prepared for measuring cognitive structures; effect size mean of reliability coefficient in adaptation studies was greater than the studies of applying and developing prepared scale. In addition to these results, it was observed that as the education levels of individuals within the sample and the number of items in scale increase, effect size mean of alpha reliability coefficient increases, too. On the other hand, it was determined that sample size and answer category number of option items did not have a direct influence on alpha coefficient. Afterwards, independent samples t test, one way ANOVA and Kruskal Wallis Test were conducted in order to find out whether the mean of Cronbach alpha acquired in line with the determined moderator variables varied according to groups. According to these analysis results, it was determined that structure of data collection tool, content of the sample, item and number of option had significant differences on Cronbach alpha coefficient. In addition to the variables within this study, different variables which are thought to affect reliability can be handled and the effects of these variables on reliability coefficient can be examined.

**Keywords:** Cronbach alpha, meta-analysis, reliability

**ÖZ:** Bu çalışmada, Hacettepe Üniversitesi Eğitim Fakültesi Dergisi'nde yayımlanan çalışmalarda yer alan Cronbach alfa güvenilirlik katsayılarından elde edilen meta analiz sonuçlarının incelenmesi amaçlanmıştır. Bu amaçla 1986-2012 yılları arasında Hacettepe Üniversitesi Eğitim Fakültesi Dergisi'nde yayınlanan 43 sayıda yer alan toplam 1222 makale incelenmiş ve dahil edilme ölçütlerine göre toplam 354 ölçme aracı ele alınmıştır. Bu araştırmada verilerin birleştirilmesi için korelasyonel verilerde etki büyüklüğü hesaplamalarındaki r indeksi; korelasyonel verilerde etki büyüklüklerinin birleştirilmesi için de Fisher z yönteminin tesadüfi etki modeli kullanılmıştır. Verilerin analizi için SPSS 20 ve MetaWin 2.0 paket programından yararlanılmıştır. Çalışmada Cronbach alfa güvenilirlik katsayısının çeşitli moderatör değişkenlere göre ortalama etki büyüklüklerinin çok güçlü olduğu sonucuna varılmıştır. Çalışma sonucunda duyuşsal yapıları ölçmek için hazırlanmış ölçeklerin güvenilirliğinin ortalama etki büyüklüğünün, bilişsel yapıları ölçmek için hazırlanmış ölçeklerinkinden; uyarlama çalışmalarındaki güvenilirlik katsayısının ortalama etki büyüklüğünün, hazır ölçeği uygulama ve geliştirme çalışmalarındakinden daha büyük olduğu belirlenmiştir. Bu sonuçların yanı sıra, örnekleme bireylerin eğitim düzeylerinin ve ölçekteki madde sayısının artmasıyla, alfa güvenilirlik katsayısının ortalama etki büyüklüğünün de arttığı gözlenmektedir. Örnekleme büyüklüğünün ve seçenek maddelerin cevap kategori sayısının ise alfa katsayısı üzerine doğrusal bir etkisi olmadığı belirlenmiştir. Daha sonra belirlenen moderatör değişkenler doğrultusunda elde edilen Cronbach alfa ortalamasının gruplara göre farklılık gösterip göstermediğini belirlemek için bağımsız örneklemler t testi, tek yönlü varyans analizi ve Kruskal Wallis testi yapılmıştır. Bu analiz sonuçlarına göre veri toplama aracının yapısının, örnekleme içeriğinin, madde ve seçenek sayısının Cronbach alfa katsayısı üzerinde anlamlı farklılıklar oluşturduğu belirlenmiştir. Bu çalışmadaki değişkenlere

\* This study was presented in "IV. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi (9-13 June 2014, Ankara".

\*\* Arş. Gör., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, e-posta: hinal@hacettepe.edu.tr

\*\*\* Arş. Gör. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, e-posta: esinyilmaz@hacettepe.edu.tr

\*\*\*\* Arş. Gör., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, e-posta: demirduzeneda@gmail.com

\*\*\*\*\* Prof. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, e-posta: gelbal@hacettepe.edu.tr

ek olarak güvenilirliği etkilediği düşünülen farklı değişkenler ele alınarak, bu değişkenlerin güvenilirlik katsayısı üzerindeki etkileri incelenebilir.

**Anahtar sözcükler:** Cronbach alfa, meta analiz, güvenilirlik

## 1. INTRODUCTION

Reliability of the results acquired from measurement tools in educational and psychological studies is one of the most important topics of the studies carried out in this area. In the widest definition, reliability can be described as clearness degree of measurement results from random errors. Up to today, lots of reliability coefficients have been recommended for estimating reliability of measurement tools. The main reason of reliability coefficients' being abundant in terms of number is that these coefficients produce different values in different item structures (Osburn, 2000). However, alpha coefficient developed by Cronbach (1951) is generally used in acquiring reliability in terms of internal consistency regarding a single test especially in combined measurements. This coefficient which is commonly used in the studies is a test reliability index (Miller, 1995) and is a coefficient calculated by using classical test theory approaches (Zumbo, 1999). This coefficient is also known as Cronbach's alpha as it was named by Cronbach in 1951 (Miller, 1995). Actually, first studies regarding coefficient were carried out by Hoyt (1941) depending on variance analysis and then, they were performed by Guttman (1945) as the sub limit of reliability (de Gruijter ve van der Kamp, 2008). Alpha coefficient depending on a single application of a scale and used in order to determine reliability can be calculated with the help of mean of the relationships between total item score variances and variances belonging to total scores (Cortina, 1993). This coefficient can be formulated as in Equation 1 as item number in  $k$  scale is variance of  $\sigma_i^2$   $i$  and variance of  $\sigma_s^2$  scale:

$$\alpha = \left( \frac{k}{k-1} \right) \left( 1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_s^2} \right) \quad \text{Equation 1}$$

Although alpha coefficient is used so frequently in the studies, the necessity of meeting some hypotheses in order to equalize alpha coefficient to actual reliability is ignored by the researchers. It is sought for normality requirement in the distribution because of ANOVA approach used in deriving alpha coefficient. Besides, additivity feature which means naturally adding actual score matrix is one of the hypotheses underlying in deriving alpha coefficient and as a result of violating this hypothesis, alpha produces values lower than actual reliability (Zimmerman, Zumbo, and Lalonde, 1993). There should not be any relationship among errors in order to calculate actual reliability of alpha coefficient (Komaroff, 1996). When this hypothesis is not met, alpha estimates reliability as higher or lower (Zumbo, 1999). Another hypothesis in using alpha coefficient is the formation of test from one dimension (Cotton, Campbell, and Malone, 1957). In his study, Tan (2009) showed that alpha internal consistency coefficients misleading researchers could be acquired when one dimensionality hypothesis was not given. In addition to these, Novick and Lewis (1967) emphasized that alpha was equal to compound measurement reliability after all components were equivalent measurements. When the alpha coefficient meets these hypotheses, it becomes equal to actual reliability.

When the studies in psychometrics are examined, it is seen that the performance of estimating alpha coefficient reliability is affected by lots of variables. Some of these factors are features of the group in which test is applied, linearity, correlations among items, dimensionality of test, distribution of item scores, narrowness of score range, number of item and option (Cortina, 1993; Duhachek, Coughlan, and Iacobucci, 2005; Enders and Bandalos, 1999; Feldt, 1993; Fife, Mendoza and Terry, 2012; Green, Lissitz, and Mulaik, 1977; Henson, 2001, Nunnally and Bernstein, 1994; Schmitt, 1996; Osburn, 2000; Tan, 2009; Tavşancıl, 2010; Yang and Green, 2011). In line with these factors specified in literature, moderator variables were chosen for the study.

Although alpha coefficient is used so frequently and it has some misusages, it is a value that is generally reported in studies. Therefore, alpha coefficient about which there are lots of studies is the focus of the study. There are many studies in which similar or different results are acquired and which are conducted independently from each other in Social Sciences. The concept of meta-analysis has come out as a result of the necessity to high level studies covering data acquired from these studies and with the help of meta-analysis, it was made saving of time, effort and cost necessary for examine the results of these studies one by one (Glass, McGaw, and Smith, 1981).

When the literature is examined, there are very few meta-analysis studies on education field in our country. On the other hand, it is seen that in most of the studies, especially in combined measurements, alpha coefficient developed by Cronbach (1951) in acquiring reliability in terms of internal consistency of a single test is frequently used. Hogan, Benjamin and Brezinski (2000) state that normative framework developed for interpreting the values acquired from Cronbach alpha coefficient which is so frequently used in the studies and the findings acquired from alpha may depend on meta-analysis. Thus, alpha coefficient which is frequently used in these studies should be analyzed comprehensively and systematically and there is also need for combined studies. On the other hand; the differences in data collection patterns such as sample size, number of item and number of option from the variables affecting performance in alpha coefficient reliability estimation; and the structure of data collection tool, preparation of data collection tool and content of the sample can be handled as variables and examining the acquired alpha coefficient is viewed as important for the field. Within the scope of this study, it was tried to answer the question of “How are the meta-analysis results that are acquired from Cronbach alpha reliability coefficients taking place in the studies published in Hacettepe University Journal of Education?” By this way, the situations among characteristics of study pattern, chosen different fictions and reliability coefficient will be able to be interpreted. Moreover, in addition to the influence quantities calculated in line with moderator variables, mean of alpha values were calculated and it was examined whether these variables had a significant difference on alpha coefficient. By this means, it is taken into account that the variables known to have effect on alpha coefficient may also create difference on mean of alpha size acquired from the combined studies.

## 2. METHOD

### 2.1. Research Model

This study which aims at obtaining information about reliability of data collection tools used in the studies taking place in Hacettepe University Journal of Education is a descriptive survey model.

### 2.2. Data Collection Tool

Within scope of this study, 1222 items taking place in 43 issues which were published in Hacettepe University Journal of Education between 1986 and 2012 were examined. On the other hand, work group of the study is 354 measurement tools reported in accordance with the criteria determined by the researchers in this study.

#### *Inclusion and Exclusion Criteria*

It is encountered with studies regarding reliability coefficient in every field. Due to the fact that reaching all studies regarding the field and including all of these studies into analysis

require so much time and effort, inclusion criteria to meta-analysis in this study were determined by the researchers as follows:

Criterion 1: The studies to be included into meta-analysis were published in Hacettepe University Journal of Education.

Criterion 2: Cronbach alpha and KR-20 reliability coefficients were calculated for reliability estimation in the study.

Criterion 3: Sample size, sample group, number of item, feature of used data collection tool (adaptation, development or adapting prepared scale) and alpha coefficient of the studies taking place in the study were included in order to be able to calculate influence quantity in meta-analysis study.

After determining inclusion criteria to the analysis, the studies not conforming to these criteria were excluded because of the fact that they were not evaluated as suitable for the study. Frequency distribution of the studies included in this study is presented in Table 1 according to years.

**Table 1: Frequency and percentage values of the studies according to years**

Variable	Categories	Frequency (f)	Percentage (%)
Year of the Study	1986-1990	6	2,76
	1991-1995	8	3,69
	1996-2000	15	6,91
	2001-2005	40	18,43
	2006-2010	84	38,71
	2011-2012	64	29,49

When Table 1 is examined, it is seen that the studies conducted between 2006 and 2010 conformed to these criteria with the rate of 38.71% at the most according to inclusion and exclusion to the study. This rate is followed by the studies conducted between 2011 and 2012 with a rate of 29.49%, 2001 and 2005 with a rate of 18.43%, 1996 and 2000 with a rate of 6.91%, 1991 and 1995 with a rate of 3.69% and 1986 and 1990 with a rate of 2.76% respectively. Due to the use of multiple measurement tools in a study, the total of 354 Cronbach alpha and KR-20 reliability coefficient was obtained from 217 article. While 80 of these coefficients are KR-20 value, the rest of 274 coefficients are Cronbach alfa value.

### 2.3. Analysis of Data

It was benefited from meta-analysis within the scope of the study. Meta-analysis is a method which combines more than one study that is independent from each other and conducted in a specific topic. It also analyzes the study findings statistically and reinterprets these findings (Lipsey and Wilson, 2001).

Phases of meta-analysis are as follows (Durlak, 1995): (1) determining research question, (2) making literature review, (3) coding studies, (4) calculating influence quantities, (5) performing statistical analyses and (6) interpreting results and reporting. These phases were followed in this study. Information regarding the first and the second steps was explained above. However, coding of the studies was carried out as follows.

#### *Coding Studies*

A coding list was formed by the researchers in order to code determined studies explicitly and comprehensively. The coding list prepared in a way that will include descriptive information and numerical results of the studies is composed of two parts. The first part is "study identity". In this part, there is information such as identity of the study, name of the study, name of the author(s) conducting the study and year of the study in which it was conducted. The second part,

however, includes “study data” such as sample quantity of the studies, item number of measurement tool and reliability values. These parts were filled for each study and the studies were coded.

### *Calculating Influence Quantities*

Influence quantities were calculated for each study included into this study. Calculation of influence quantity was made in correlational data due to the fact that the relationship of Cronbach alpha coefficient with different variables was examined in this study; therefore it was benefited from r index as influence quantity index.

### *Conducting Statistical Analyses*

Fisher z method was used in this phase in order to combine influence quantities acquired from correlational data. Fisher z conversion was carried out so as to be able to make more sensitive estimations in correlation data.

However, in order to combine influence quantities,

- heterogeneity of influence quantities should be acceptable and
- influence quantities should show normal distribution.

Thus, these hypotheses were tested before dealing with statistical analyses. In this study, it was benefited from Q statistics in order to determine heterogeneity. Significance level of meta-analysis study is recommended 0.10 (Hedges and Olkin, 1985). When the results acquired from heterogeneity test are examined, it is seen that total heterogeneity value is significant ( $Q_{Total(353)} = 2455.556, p < 0.10$ ). Getting significant results may show that some moderator variables should be handled as well as it may be a result of partiality. In these kinds of situations, random influence model can be used and this model was used in this study.

It can be calculated how many studies should be included into the analysis that will give the result of zero regarding influence quantity about the topic in order to resolve the problem of partiality of the studies included into meta-analysis studies. Calculated number is called fail safe number (Rosenberg, 2005). In this study, fail safe number was found as 65435282,2 according to Rosenthal method and as 1592,4 according to Orwin method. That is, when it is considered even for Orwin Method, mean meta-analysis result decreases to 0,05 when 1592 more studies whose influence quantity value is 0 are included into meta-analysis. When it is considered that meta-analysis is conducted in 354 measurement tools, it can be said that 1592 studies possess high values and accordingly, meta-analysis results are reliable. In this way, it can be said that heterogeneity test's being significant is not a result of a partiality. However, heterogeneity test's being significant may show that it should be dealt with moderator variables. Thus, this study was conducted under different moderator variables. Moderator variables of the study was determined as structure of data collection tool, preparation of data collection tool, sample content, sample size and number of options. Furthermore, the conformity of influence quantities to normal distribution was examined and it was decided that there was not any problem.

The classification performed by Cohen, Manion and Morrison (2007) was used while interpreting the significance of influence quantities in the study. According to this classification, influence quantities are classified as,

- weak between 0,00 and  $\pm 0,10$ ,
- small between  $\pm 0,10$  and 0,30,
- moderate between  $\pm 0,30$  and 0,50,
- strong between  $\pm 0,50$  and 0,80,
- very strong over  $\pm 0,80$ .

Independent samples t test and one way variance analysis were carried out in order to find out whether Cronbach alpha coefficient showed significant difference according to sub-groups of

determined moderator variables. Also, due to the small number of data in some group of option number, Kruskal Wallis Test was carried out in order to find out whether Cronbach alpha coefficient showed significant difference according to number of option. The Mann-Whitney U test was conducted to determine if there are differences between which groups. It was benefited from MetaWin 2.0 packet program for the analysis of data.

### 3. FINDINGS

This chapter explains the step of interpreting and reporting results which is the final phase of meta-analysis. In this study which is conducted in accordance with 354 measurement tools published in Hacettepe University Journal of Education, meta-analysis of Cronbach alpha reliability coefficient was performed from different moderator variables. Effect size mean and 95% confidence intervals of Cronbach alpha reliability coefficient are given in Table 2 according to the structure of data collection tool, preparation of data collection tool, type of sample, sample size, number of items and options.

**Tablo 2: The results of effect size mean according to different moderator variables**

Moderator Variable	Category	Number of Work	Effect Size Mean	%95 Confidence Intervals for Effect Size	
				Lower Limit	Upper Limit
Structure of Data Collection Tool	Affective	212	1,386	1,368	1,403
	Cognitive	142	1,112	1,090	1,135
Preparation of Data Collection Tool	Prepared	107	1,252	1,227	1,277
	Development	196	1,277	1,257	1,298
	Adaptation	51	1,361	1,323	1,400
Type of Sample	Primary School	95	1,105	1,079	1,132
	High School	69	1,243	1,212	1,275
	Higher Education	138	1,344	1,321	1,366
	Teacher	28	1,545	1,495	1,595
	Other	24	1,433	1,381	1,486
Sample Size	≤50	43	1,326	1,260	1,392
	>50 ve ≤100	70	1,201	1,161	1,240
	>100 ve ≤200	91	1,297	1,267	1,328
	>200	149	1,286	1,265	1,306
Number of Items	≤10	29	1,092	1,039	1,145
	>10 ve ≤20	104	1,198	1,170	1,227
	>20 ve ≤30	102	1,256	1,231	1,282
	>30	93	1,477	1,449	1,505
	Not given	26	1,195	1,143	1,247
Number of Option	Not given	72	1,210	1,177	1,242
	2	82	1,124	1,095	1,153
	3	11	1,490	1,406	1,573
	4	29	1,081	1,030	1,132
	5	142	1,401	1,380	1,423
	6	6	1,360	1,230	1,491
	7	10	1,455	1,370	1,539
	8	2	1,653	0,411	2,895
	9	2	1,653	0,411	2,895

There are effect sizes mean according to various moderator variables of 354 Cronbach alpha reliability coefficients in the analysis and 95% confidence intervals in Table 2. When the effect sizes are interpreted in the study, it can be said that all of the effect sizes in the table are very strong according to the classification of Cohen et al. (2007).

When the alpha coefficient was examined according to the structure of data collection tool, effect size mean of the reliability acquired from scales prepared for measuring affective structures was found as 1,386 while lower limit was calculated as 1,368 and upper limit as 1,403

in 95% confidence interval. It was determined that, effect size mean of the reliability acquired from scales prepared for measuring cognitive structures was 1,112 while lower limit was calculated as 1,090 and upper limit as 1,135 in 95% confidence interval. In this situation, it can be said that, effect size mean of the scales' reliability prepared for measuring affective structures is greater than the scales prepared for measuring cognitive structures. Moreover, when the effect size mean is examined, it can be said that structures of the both data collection tools have positive and great effect on alpha reliability coefficient estimation.

Effect size mean of reliability coefficient for the studies in which prepared scale used was 1,252; lower limit was calculated as 1,227 and upper limit as 1,277 in 95% confidence interval. Effect size mean in development studies was 1,277, lower limit was determined as 1,257 and upper limit as 1,298 in 95% confidence interval. On the other hand, effect size mean in adaptation studies was 1,361, lower limit was calculated as 1,323 and upper limit as 1,400 in 95% confidence interval. According to these findings, it can be said that effect size mean of reliability coefficient in adaptation studies has greater values than prepared scale and development studies. In addition to this, it can be said that the way of preparing data collection tool has a positive and great effect on alpha reliability coefficient estimation.

Effect size mean sample content of reliability coefficient is 1,105 when it is for primary school students; lower limit is 1,079 and upper limit is 1,132 in 95% confidence interval. Effect size mean sample content of reliability coefficient is 1,243 when it is for high school students; lower limit is 1,212 and upper limit is 1,275 in 95% confidence interval. Effect size mean sample content of reliability coefficient is 1,344 when it is for higher education students; lower limit is 1,321 and upper limit is 1,366 in 95% confidence interval. Effect size mean sample content of reliability coefficient is 1,545 for teachers; lower limit is 1,495 and upper limit is 1,595 in 95% confidence interval. Effect size mean sample content of reliability coefficient is 1,433 for school principals; lower limit is 1,381 and upper limit is 1,486 in 95% confidence interval. According to these results, it can be said that as the education level of individuals within the sample increases, effect size mean of alpha reliability coefficient increases, too. In addition, it can be said that sample content has a positive and great effect on alpha reliability coefficient estimation in all situations.

Effect size mean of reliability coefficient is 1,326 when the sample size is 50 and less; lower limit is 1,260 and upper limit is 1,392 in 95% confidence interval. Effect size mean influence quantity of reliability coefficient is 1,201 when the sample size is between 50 and 100; lower limit is 1,161 and upper limit is 1,240 in 95% confidence interval. Effect size mean belonging to alpha coefficient is 1,297 when the sample size is 100 and 200; lower limit is 1,267 and upper limit is 1,328 in 95% confidence interval. Lastly, effect size mean value belonging to alpha coefficient is 1,286 when the sample size is 200 and over; lower limit is 1,265 and upper limit is 1,306 in 95% confidence interval. In addition to these findings, it can be stated alpha all alpha estimations of sample sizes have positive and great effect on effect size mean.

Effect size mean of reliability coefficient is 1,091 when the number of item is 10 and less; lower limit is 1,039 and upper limit is 1,145 in 95% confidence interval. Effect size mean of alpha coefficient is 1,198 when the number of item is between 10 and 20; lower limit is 1,170 and upper limit is 1,227 in 95% confidence interval. Effect size mean of alpha coefficient is 1,256 when the number of item is between 20 and 30; lower limit is 1,231 and upper limit is 1,282 in 95% confidence interval. Effect size mean of alpha coefficient is 1,477 when the number of item is between 30 and over; and it is seen that this value is greater than the values acquired from other situations. When the number of item is more than 30, lower limit is 1,449 and upper limit is 1,505 in 95% confidence interval regarding effect size mean. Moreover, when

the effect size mean is examined, it can be said that the number of item has a positive and great effect on alpha reliability coefficient estimation in all situations.

Effect size mean of reliability coefficient is 1,12 when the number of option is two; lower limit is 1,095 and upper limit is 1,153 in 95% confidence interval. Effect size mean of alpha coefficient is 1,490 when the number of option is three; lower limit is 1,406 and upper limit is 1,573 in 95% confidence interval. Effect size mean of alpha coefficient is 1,081 when the number of option is four; lower limit is 1,030 and upper limit is 1,132 in 95% confidence interval. When the number of option is five; lower limit is 1,170 and upper limit is 1,227 in 95% confidence interval. Effect size mean of alpha coefficient is 1,360 when the number of option is six; lower limit is 1,230 and upper limit is 1,491 in 95% confidence interval. Effect size mean of alpha coefficient is 1,455 when the number of option is seven; lower limit is 1,370 and upper limit is 1,539 in 95% confidence interval. Lastly, effect size mean of alpha coefficient is 1,653 when the number of option is nine; lower limit is 0,411 and upper limit is 2,895 in 95% confidence interval. It can be stated that the number of options has a positive and great effect on alpha reliability coefficient estimation in all situations.

Independent samples t test and one way variance analysis was carried out in order to determine whether alpha coefficients acquired from all studies varied according to moderator variables by taking mean of these coefficients. t test results to have been carried out in order to examine difference situation of alpha values according to the structure of data collection tool are presented in Table 3.

**Table 3: The results of t test according to the structure of data collection tool**

Moderator Variable	Category	Number of Work	Mean	Std. Deviation	sd	t	p
The Structure of Data	Affective	212	0,86	0,871	352	7,313	,000
Collection Tool	Cognitive	142	0,79	0,874			

When Table 3 is examined, it is seen that mean of alpha value acquired from measurement tools measuring affective structures is 0,86 and mean of alpha value acquired from measurement tools measuring cognitive structures is 0,79. Moreover, it was determined that mean of Cronbach alpha reliability value showed a significant difference according to the structure of data collection tool ( $t_{(352)} = 7.313$ ,  $p < .01$ ). According to these findings, it can be said that affective data collection tools have higher alpha values than cognitive data collection tools.

One way variance analysis was performed in order to find out whether Cronbach alpha reliability coefficient of 354 measurement tools included into the analysis varied according to various moderator variables. ANOVA results are shown in Table 4. When Table 4 is examined in which results regarding whether reliability coefficient mean varied significantly according to sub-groups of moderator variables, it could be seen that mean of alpha value was 0,83 which was acquired from the studies in which prepared scales were used or scales were developed while mean of alpha value was 0,85 which was acquired from the adapted scales. Furthermore, it was found out that the mean of alpha values acquired according to the structure of data collection tool were not so variable and this variability did not make a significant difference ( $p > .01$ ).

When the mean value of reliability coefficient was examined according to sample content, the highest value was taken with 0,88 if the mean of alpha value sample content was teacher; whereas the lowest value was taken as 0.79 if the mean of alpha value sample content was primary school students. These differences among mean of alpha values were found significant ( $p < .01$ ). It was determined in Tukey test which was conducted in order to find out among which group significant difference was that the group whose age level was high showed a great difference when compared to primary school group.



**Table 4: The results of ANOVA according to various moderator variable**

Moderator Variable	Category	Number of Work	Mean	F	p	Significant Difference
Preparation of Data Collection Tool	Prepared	107	0,83	1,301	,274	-
	Development	196	0,83			
	Adaptation	51	0,85			
Type of Sample	Primary School (1)	95	0,79	8,070	,000	1-3, 1-4, 1-5
	High School (2)	69	0,82			
	Higher Education (3)	138	0,84			
	Teacher (4)	28	0,88			
	Other (5)	24	0,87			
Sample Size	≤50	43	0,83	,346	,792	-
	>50 ve ≤100	70	0,81			
	>100 ve ≤200	91	0,84			
	>200	149	0,83			
Number of Items	≤10 (1)	29	0,78	11,701	,000	0-4, 1-3, 1-4, 2-3, 2-4, 3-4
	>10 ve ≤20 (2)	104	0,80			
	>20 ve ≤30 (3)	102	0,83			
	>30 (4)	93	0,87			
	Not given (0)	26	0,82			

Close results were acquired when the mean of alpha values were examined according to different sample sizes and it was determined that there was not a significant difference among these values ( $p < .01$ ). Accordingly, it can be said that sample size did not have a significant effect on alpha coefficient quantity.

Mean of value of reliability coefficient was 0,78 when the number of item was 10 and lower; 0,80 between 10 and 20; 0,83 between 20 and 30 and 0,87 when the number of item was 30 and higher. Therefore, it can be said that the increase in the number of items in measurement tool might increase reliability coefficient, too. Furthermore, it was determined that the increase in the number of items showed a significant difference on alpha estimation ( $p < .01$ ). According to Post Hoc test, acquired alpha value when the number of items was generally high showed difference compared to the acquired alpha value when the number of items was generally low.

Kruskal Wallis Test was carried out in order to find out whether Cronbach alpha coefficient showed significant difference according to number of options. The Mann-Whitney U test was conducted to determine if there are differences between which groups. Results are shown in Table 5.

**Table 5: The result of Kruskal Wallis test according to number of option**

Moderator Variable	Category	Number of Work	Mean Rank	X <sup>2</sup>	p	Significant Difference
Number of Option	2	82	102,46	47,938	,000	2-3, 2-5, 2-4, 2-6, 2-7, 2-9, 3-4, 4-5, 4-6, 4-7, 4-9
	3	11	170,95			
	4	29	96,12			
	5	142	165,20			
	6	6	171,50			
	7	10	187,25			
	9	2	236,50			

When Table 5 is examined, it was seen that the highest value (236.50) of mean rank of alpha occurred when the number of options was nine. It was found that there was a significant difference among alpha mean values calculated according to number of options ( $p < .01$ ). It was observed in the paired comparisons that generally when the alpha value acquired from situations in which number of options was high showed difference compared to the situations in which

number of options was low. In research articles, the findings should be given here and the above mentioned principles should be considered.

#### 4. DISCUSSION and RESULTS

In this study, it was concluded that effect size mean was very strong according to various moderator variables of Cronbach alpha reliability coefficient. Besides, effect size mean of reliability of scales prepared for measuring affective structures was greater than scales prepared for measuring cognitive structures; effect size mean of reliability coefficients in adaptation studies was greater than those prepared for application and development of prepared scales. In addition to these results, it is observed that as the education level of individuals within the sample increases, effect size mean of alpha reliability coefficient increases, too. Similar to the results of this study, Peterson (1994) determined that sample size did not have a significant effect on alpha coefficient. Similar to the results of Churchill and Peter's (1984) study, it is seen in this study that the number of items has a significant effect on the effect size mean of reliability coefficient. In contrast to this study, Churchill and Peter (1984) as well as Peterson (1994) determined that the number of options belonging to the items had an effect on alpha coefficient quantity. All of influence quantities calculated in terms of all moderator variables were positive and in great level and this might result from researchers' avoiding from publishing their studies possessing low alpha values.

In this study, the means of alpha coefficients were also calculated according to sub-groups of variables determined as moderator and that mean also examined whether alpha values varied according to sub-groups. When the means of alpha coefficient values were calculated separately according to the structure of data collection tool, it was determined that the mean of alpha value (0,86) acquired from measurement tools measuring affective structures was higher than mean of alpha value (0,79) acquired from measurement tools measuring cognitive structures and there was a significant difference among those values. Although reliability limit determined for tests which measured affective features was lower than reliability limit determined for tests which measured cognitive features, it was determined in the analyses performed in this study that affective tests were more reliable. However, both of the mean of reliability values acquired according to the structure of data collection tool were found to be moderately high according to Murphy and Davidshofer's (1988) classification.

When the means of alpha values were examined according to the preparation of data collection tool, it was determined that mean of alpha value acquired from the adapted scales was higher than mean of alpha value acquired from the studies in which prepared scales were used and scales were developed. However, it was not found a significant difference among the means of alpha values acquired according to the structure of data collection tool. It is thought that this situation was resulted from the fact that researchers might mostly want to reach reliable measurement tools independent from the situation of used measurement tool.

When the results are examined in terms of sample content of the mean value of reliability coefficient, it is seen that as age level increases, the mean of alpha value increases, too. The differences resulting from sample content among mean of alpha values were found significant. The reason of this might be the fact that as age level increases, individuals responded measurement tools more sincerely and seriously.

When the means of alpha values were examined according to different sample sizes, similar results were acquired and it was determined that there was not a significant difference among those values. Therefore, it can be said that sample size did not have a significant effect on

alpha coefficient size. There are some views in the literature that sample size should generally be high in reliability estimations (Charter, 1999, 2003; Kline, 1986, Nunnally and Bernstein, 1994; Segall, 1994). However, Yurdugül (2008) specifies in his/her study that an impartial and coherent alpha estimation depends on size of first eigenvalue as well as size of sample width and even in in 30 sample size, when the first eigenvalue is  $\geq 10$ , an impartial estimation of alpha coefficient can be made. The result acquired in this study regarding sample size may stem from a situation about eigenvalue volume.

Findings acquired by examining mean value of reliability coefficient according the number of item show that the increase of number of items in measurement tool may also increase reliability coefficient. Furthermore, it was determined that the increase of item number had a significant difference on alpha estimation. This result shows parallelism with the study results determining that article number will increase alpha estimation (Cortina, 1993, Green et al. 1977). However, it was shown that even when item number of alpha coefficient in a multi-dimensional measurement whose inter-item correlations were so low was provided, a reliability to be evaluated as pretty high and acceptable by most of the standards (0.70 and higher) might be acquired (Cortina, 1993). This is because although alpha coefficient is so sensitive to internal consistency, that is, to the relationships among items, it is affected by the number of items in the scale so frequently. Thus, in order to make it possible to calculate alpha coefficient equal to actual reliability, it may be suggested for the researchers to make sure that alpha coefficient meets basic hypotheses instead of increasing number of items.

Lastly, when it was observed that the change of the number of options has a significant difference on rank mean of alpha value, it was determined that there was a significant difference among means of alpha quantities calculated according to the number of options. If the results are handled generally, alpha value acquired from the situations in which the number of options was high was calculated as higher than alpha value acquired from the situations in which number of option was low. It is shown in Enders and Bandalos's (1999) study that the number of options affected alpha coefficient estimation. In this study, similar results as in the study of Enders and Bandalos (1999) were acquired in terms of number of option and alpha estimation.

In this study, only those articles in Hacettepe University Journal of Education were examined. A more comprehensive study may be conducted by using thesis, bulletin, and other studies and different journals, publications which calculate reliability with Cronbach alpha coefficient. Moreover, different variables which are thought to affect reliability and their effects can be examined by taking the moderator variable. The effect of moderator variables among which there are significant differences on alpha can be handled with a comprehensive study. It is suggested that meta-analysis studies conducted in education field should be increased in our country. Consequently, this study may be repeated by taking the studies of researchers into consideration regarding Cronbach alpha coefficient which is frequently used.

## 5. REFERENCES

- Charter, R. A. (1999). Sample size requirements for precise estimates of reliability, generalizability, and validity coefficients. *Journal of Clinical and Experimental Neuropsychology*, 21, 559-566.
- Charter, R.A. (2003). Study samples are too small to produce sufficiently precise reliability coefficients. *The Journal of General Psychology*, 130, 117-129.
- Churchill, G. A., & Peter, J. P. (1984). Research design effects on the reliability of rating scales: A meta-analysis. *Journal of Marketing Research*, 21(4), 360-375.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (Sixth Edition). London and New York: Rotledge.
- Cortina, J. M (1993). What is coefficient alpha? An examination of theory and application. *Journal of Applied Psychology*, 78, 98-104.

- Cotton, J. W., Campbell, D. T., & Malone, R. D. (1957). The relationship between factorial composition of test items and measures of test reliability. *Psychometrika*, 22(4), 347-357.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- de Gruijter, D. N. M., & van der Kamp, L. J. T. (2008). *Statistical test theory for the behavioral sciences*. Boca Raton, FL: Chapman & Hall / CRC.
- Duhachek, A., Coughlan, A. T., & Iacobucci, D. (2005). Results on the standard error of the coefficient alpha index of reliability. *Marketing Science*, 24(2), 294-301.
- Durlak, J. A. (1995). *Reading and understanding multivariate statistics*. Washington DC: American Psychological Association.
- Enders, C. K., & Bandalos, D. L. (1999). The effects of heterogeneous item distributions on reliability. *Applied Measurement in Education*, 12(2), 133-150.
- Feldt, L. S. (1993). The relationship between the distribution of item difficulties and test reliability. *Applied Measurement in Education*, 6, 37-48.
- Fife, D. A., Mendoza, J. L., & Terry, R. (2012). The assessment of reliability under range restriction: a comparison of  $\alpha$ ,  $\omega$ , and test-retest reliability for dichotomous data. *Educational and Psychological Measurement*, 20(10), 1-27.
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage Publications.
- Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, 37, 827-838.
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10, 255-282.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
- Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development*, 34(3), 177-189.
- Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods. *Educational and Psychological Measurement*, 60, 523-531.
- Hoyt, C. (1941). Test reliability estimated by analysis of variance. *Psychometrika*, 6, 153-160.
- Kline, P. (1986). *A handbook of test construction: Introduction to psychometric design*. New York: Methuen.
- Komaroff, E. (1996). *Coefficient alpha under simultaneous violations of essential tau-equivalence and uncorrelated errors*. Doctoral dissertation, University of Miami, FL.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. California: Sage Publication.
- Miller, M. B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. *Structural Equation Modeling*, 2(3), 255-273.
- Murphy, K. R., & Davidshofer, C. O. (1988). *Psychological testing: Principles and application*. Englewood Cliffs, NJ: Prentice-Hall.
- Novick, M. R., & Lewis, C. L. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, 32, 1-13.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd Ed.). New York: McGraw-Hill.
- Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods*, 5(3), 343-335.
- Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research*, 21(2), 381-39.
- Rosenberg, M. S. (2005). The file-drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution*, 59(2), 464-468.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8, 350-353.
- Segall, D. O. (1994). The reliability of linearly equated tests. *Psychometrika*, 59, 361-375.
- Tan, Ş. (2009). Misuses of KR-20 and Cronbach's alpha reliability coefficients. *Education and Science*, 34(152), 101-112.
- Tavşancıl, E. (2010). *Tutumların ölçülmesi ve SPSS ile veri analizi*, (4. Baskı), Ankara: Nobel Yayın Dağıtım.
- Yang, Y., & Green, S. B. (2011). Coefficient Alpha: A Reliability Coefficient for the 21st Century? *Journal of Psychoeducational Assessment*, 29, 377-392.
- Yurdugul, H. (2008). Minimum sample size for Cronbach's coefficient alpha: A monte-carlo study. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 35, 397-405.
- Zimmerman, D. W., Zumbo, R. D., & Lalonde, C. (1993). Coefficient alpha as an estimate of test reliability under violation of two assumptions. *Educational and Psychological Measurement*, 53, 33-49.

Zumbo, B. D. (1999). *A glance at coefficient alpha with an eye towards robustness studies: Some mathematical notes and a simulation model*. Edgeworth Series in Quantitative Behavioral Science, UNBC. Retrieved January 2012, <http://www.educ.ubc.ca/faculty/zumbo/alpha99-1.pdf>

## Uzun Özet

Eğitimsel ve psikolojik araştırmalarda ölçme araçlarından elde edilen sonuçların güvenilirliği bu alanda yapılan çalışmaların en önemli konularından birisidir. En genel tanımıyla güvenilirlik, ölçme sonuçlarının tesadüfi hatalardan arınlık derecesi olarak ifade edilebilir. Günümüze kadar ölçme araçlarının güvenilirliğini kestirmeye yönelik çok sayıda güvenilirlik katsayısı önerilmiştir. Güvenirlik katsayılarının sayıca çok olmasının temel nedeni; bu katsayıların farklı madde yapılarında farklı değer üretmeleridir (Osburn, 2000). Ancak, özellikle birleşik ölçmelerde tek bir teste ilişkin iç tutarlılık anlamındaki güvenilirliğin elde edilmesinde Cronbach (1951) tarafından geliştirilen alfa katsayısının yaygın bir şekilde kullanılmaktadır. Çalışmalarda çok yaygın olarak kullanılan bu katsayı, bir test güvenilirliği indeksidir (Miller, 1995) ve klasik test teorisi yaklaşımlarını kullanarak hesaplanan bir katsayıdır (Zumbo, 1999). Bu katsayı 1951 yılında Cronbach tarafından isimlendirildiği için Cronbach'ın alfası olarak da bilinir (Miller, 1995). Aslında katsayı ile ilgili ilk çalışmalar varyans analizi temeline dayanarak Hoyt (1941) tarafından, daha sonra güvenilirliğin alt sınırı olarak Guttman (1945) tarafından yapılmıştır (de Gruijter ve van der Kamp, 2008). Bir ölçeğin tek bir uygulamasına dayanan ve güvenilirliği belirlemek için kullanılan alfa katsayısı, madde puanları varyansları toplamı ile toplam puanlara ait varyanslar arasındaki ilişkilerin ortalaması yardımıyla hesaplanabilmektedir (Cortina, 1993).

Araştırmalarda alfa katsayısının bu kadar sık kullanılmasına karşın, alfa katsayısının gerçek güvenilirliğe eşit olması için bazı varsayımları karşılaması gerektiği araştırmacılar tarafından göz ardı edilmektedir. Alfa katsayısının türetilmesinde kullanılan ANOVA yaklaşımı nedeniyle dağılımda normallik şartı aranır. Ayrıca gerçek puanlar matrisinin doğal olarak toplanması anlamına gelen toplanabilirlik özelliği de alfa katsayısının türetilmesi altında yatan varsayımlardandır ve bu varsayımın ihlali sonucunda alfa, gerçek güvenilirliğin altında değerler üretmektedir (Zimmerman, Zumbo ve Lalonde, 1993). Alfa katsayısının gerçek güvenilirliği hesaplayabilmesi için hatalar arasında bir ilişki olmaması da gerekmektedir (Komaroff, 1996). Bu varsayım karşılanmadığında ise alfa, güvenilirliği daha yüksek ya da daha düşük tahmin eder (Zumbo, 1999). Alfa katsayısının kullanımındaki bir diğer varsayım ise testin tek boyuttan oluşmasıdır (Cotton, Campbell ve Malone, 1957). Tan (2009) yaptığı araştırmada, tek boyutluluk varsayımının sağlanmadığı durumlarda araştırmacıları yanlış yönlendiren alfa iç tutarlılık katsayılarının elde edilebileceği göstermiştir. Bunların yanı sıra Novick ve Lewis (1967) de bütün bileşenlerin eşdeğer ölçmeler olması sonucunda alfanın, bileşik ölçmelerin güvenilirliğine eşit olduğunu vurgulamıştır. Alfa katsayısının bu varsayımları karşıladığında gerçek güvenilirliğe eşit olmaktadır.

Alfa katsayısının çok sık ve bazı yanlış kullanımları olsa da, araştırmalarda genellikle rapor edilen bir değerdir. Üzerinde bu kadar çok çalışma bulunan alfa katsayısı bu nedenle çalışmanın odağını oluşturmuştur. Sosyal Bilimler alanında, aynı konuda birbirinden bağımsız yapılan ve benzer ya da farklı sonuçlara ulaşılan birçok çalışma bulunmaktadır. Bu çalışmalardan elde edilen bilgileri kapsayacak üst çalışmalara olan gereklilik sonucunda meta analiz kavramı ortaya çıkmıştır ve meta analiz yardımıyla da bu çalışmaların sonuçlarını tek tek incelemek için gereken zamandan, çabadan ve maliyetten tasarruf sağlanmıştır (Glass, McGaw ve Smith, 1981).

Alan yazın incelendiğinde ülkemizde eğitim alanında çok az sayıda meta-analiz çalışmasına rastlanılmıştır. Diğer yandan araştırmaların çoğunda, özellikle birleşik ölçmelerde tek bir teste ilişkin iç tutarlılık anlamındaki güvenilirliğin elde edilmesinde Cronbach (1951) tarafından geliştirilen alfa katsayısının yaygın bir şekilde kullanıldığı görülmektedir. Hogan, Benjamin ve Brezinski (2000) araştırmalarda bu kadar çok kullanılan Cronbach alfa katsayısından elde edilen değerleri yorumlamak için geliştirilen normatif çerçevenin, alfadan elde edilen bulguların meta analize dayanıyor olabileceğini belirtmişlerdir. Bu nedenle de araştırmalarda çok sık kullanılan alfa katsayısını kapsamlı ve sistematik bir şekilde analiz ederek birleştiren araştırmalara ihtiyaç duyulmaktadır. Diğer yandan alfa katsayısının güvenilirliği kestirimindeki performansını etkileyen değişkenlerden örneklem büyüklüğü, madde sayısı ve seçenek sayısı; ayrıca veri toplama aracının yapısı, veri toplama aracının hazırlanışı ve örneklem içeriği şeklindeki araştırma desenindeki farklılıklar değişken olarak ele alınarak, elde edilen alfa katsayılarını incelemek de alan için önemli görülmektedir. Bu araştırma kapsamında "Hacettepe Üniversitesi Eğitim

Fakültesi Dergisi'nde yayınlanan çalışmalarda yer alan Cronbach alfa güvenilirlik katsayılarından elde edilen meta analiz sonuçları nasıldır?" sorusuna yanıt aranmaya çalışılmıştır.

Hacettepe Üniversitesi Eğitim Fakültesi Dergisi'nde yer alan çalışmalarda kullanılan veri toplama araçlarının güvenilirliğine ilişkin bilgi edinmeyi hedefleyen bu araştırma betimsel bir araştırma olup tarama modelindedir. Bu araştırma kapsamında, 1986-2012 yılları arasında Hacettepe Üniversitesi Eğitim Fakültesi Dergisi'nde yayınlanan 43 sayıda yer alan toplam 1222 makale incelenmiştir. Araştırmanın çalışma grubunu ise bu makalelerde araştırmacılar tarafından belirlenen ölçütlere uygun olarak rapor edilmiş 354 ölçme aracı oluşturmaktadır.

Bu çalışmada Cronbach alfa güvenilirlik katsayısının çeşitli moderatör değişkenlere göre ortalama etki büyüklüklerinin çok güçlü olduğu sonucuna varılmıştır. Ayrıca duyuşsal yapıları ölçmek için hazırlanmış ölçeklerin güvenilirliğinin ortalama etki büyüklüğünün, bilişsel yapıları ölçmek için hazırlanmış ölçeklerinkinden; uyarlama çalışmalarındaki güvenilirlik katsayılarının ortalama etki büyüklüğünün, hazır ölçeği uygulama ve geliştirme çalışmalarındakinden daha büyük olduğu belirlenmiştir. Bu sonuçların yanı sıra, örnekleme bireylerin eğitim düzeylerinin artmasıyla, alfa güvenilirlik katsayısının ortalama etki büyüklüğünün de arttığı gözlenmektedir. Peterson (1994), bu çalışmadaki sonuçlara benzer olarak örneklem büyüklüğünün alfa katsayısı üzerine anlamlı bir etkisi olmadığını belirlemiştir. Bu çalışmada Churchill ve Peter'in (1984) araştırmasındaki sonuçlara benzer olarak, madde sayısının güvenilirlik katsayısının ortalama etki büyüklüğüne anlamlı bir etkisi olduğu görülmektedir. Bu çalışmanın aksine Churchill ve Peter'in (1984) ile Peterson (1994) maddelere ait seçenek sayısının ise alfa katsayısının büyüklüğü üzerinde etkili olduğunu belirlemişlerdir. Bütün moderatör değişkenler açısından hesaplanan etki büyüklüklerinin tamamının pozitif yönde ve büyük düzeyde çıkması, araştırmacıların düşük alfa değerine sahip olan çalışmaları yayınlamaktan kaçınmalarından kaynaklanmış olabileceğini akla getirmektedir.

Araştırmada moderatör olarak belirlenen değişkenlerin alt gruplarına göre ortalama alfa katsayıları da hesaplanmış ve bu ortalama alfa değerlerinin alt gruplara göre farklılık gösterip göstermediği de incelenmiştir. Alfa katsayısının ortalamasının veri toplama aracının yapısına göre değerleri ayrı ayrı hesaplandığında duyuşsal yapıları ölçen ölçme araçlarından elde edilen ortalama alfa değerinin (0,86) ve bilişsel yapıları ölçen ölçme araçlarından elde edilen ortalama alfa değerinden (0,79) büyük olduğu ve bu değerler arasında anlamlı bir farklılık olduğu belirlenmiştir. Duyuşsal özellikleri ölçen testler için belirlenen güvenilirlik sınırı, bilişsel özellikleri ölçen testlerin güvenilirlik sınırından daha düşük olsa da, bu çalışmada yapılan analizler doğrultusunda duyuşsal testlerin daha güvenilir olduğu belirlenmiştir. Ancak veri toplama aracının yapısına göre elde edilen her iki ortalama güvenilirlik değeri de Murphy ve Davidshofer'in (1988) sınıflamasına göre orta düzeyde yüksek bulunmuştur.

Veri toplama aracının hazırlanışına göre ortalama alfa değerleri incelendiğinde, uyarlama yapılan ölçeklerden elde edilen ortalama alfa değerinin, hazır ölçeğin kullanıldığı ve ölçeğin geliştirildiği çalışmalardan elde edilen ortalama alfa değerinden büyük olduğu belirlenmiştir. Ancak veri toplama aracının yapısına göre elde edilen ortalama alfa değerleri arasında anlamlı farklılık bulunmamıştır. Bu durum kullanılan ölçme aracının durumundan bağımsız olarak, araştırmaların çoğunlukla güvenilir ölçme araçlarına ulaşmak istemelerinden kaynaklanıyor olabileceği düşünülmektedir.

Sonuçlar güvenilirlik katsayısının ortalama değerinin örneklem içeriği bakımından incelendiğinde, yaş düzeyi arttıkça ortalama alfa değerinin de arttığını göstermektedir. Ortalama alfa değerleri arasında örneklem içeriğinden kaynaklanan farklılıklar anlamlı bulunmuştur. Bu durum yaş düzeyi arttıkça, bireylerin ölçme araçlarını daha samimiyetle ve ciddiyetle yanıtlaması nedeniyle meydana gelmiş olabilir.

Ortalama alfa değerleri de farklı örneklem büyüklüklerine göre incelendiğinde yakın sonuçlar elde edilmiştir ve bu değerler arasında anlamlı farklılık oluşmadığı belirlenmiştir. Bu nedenle, örneklem büyüklüğünün alfa katsayısının büyüklüğü üzerinde anlamlı bir etkisi olmadığı söylenebilir. Alan yazında güvenilirlik kestirimlerinde örneklem büyüklüğünün genel olarak fazla olması gerektiği yönünde görüşler bulunmaktadır (Charter, 1999, 2003; Kline, 1986, Nunnally ve Bernstein, 1994, Segall, 1994). Ancak Yurdugül (2008) yaptığı çalışmada yansız ve tutarlı bir alfa kestirimi örneklem genişliğinin büyüklüğü kadar aynı zamanda ölçmelerin birinci özdeğerinin büyüklüğüne bağlı olduğunu, 30 örneklem büyüklüğünde bile birinci özdeğer  $\geq 10$  olduğunda alfa katsayısının yansız bir kestirimini yapılabileceğini belirtmiştir. Bu araştırmadaki örneklem büyüklüğü ile ilgili ulaşılan sonuç, özdeğer yükleri ile ilgili bir durumdan kaynaklanıyor olabilir.

Güvenirlilik katsayısının ortalama değeri madde sayısına göre incelenmesiyle elde edilen bulgular, ölçme aracındaki madde sayısının artmasının güvenirlilik katsayısını da arttırmış olabileceği göstermektedir. Ayrıca madde sayısının artmasının alfa kestirimi üzerinde anlamlı bir fark yarattığı belirlenmiştir. Son olarak seçenek sayısının değişmesinin alfa değerinin sıra ortalaması üzerinde anlamlı farklılık yaratıp yaratmadığına bakıldığında, seçenek sayısına göre hesaplanan sıra ortalamaları arasında anlamlı farklılık olduğu belirlenmiştir.

Bu çalışmada yalnızca Hacettepe Üniversitesi Eğitim Fakültesi Dergisi'ndeki makaleler incelenmiştir. Güvenirliliği Cronbach alfa katsayısı ile hesaplayan tez, bildiri vb. gibi farklı türdeki araştırmalar ve farklı dergiler, yayınlar kullanılarak daha kapsamlı bir çalışma yapılabilir. Ayrıca güvenirliliği etkilediği düşünülen farklı değişkenler, moderatör değişken alınarak, bu değişkenlerin etkileri incelenebilir. Aralarında anlamlı fark bulunan moderatör değişkenlerin, alfa üzerindeki etkisi kapsamlı bir çalışma ile ele alınabilir. Eğitim alanında yapılan meta analiz çalışmalarının ülkemizde daha da artması tavsiye edilebilir. Son olarak araştırmalarda bu kadar sık kullanılan Cronbach alfa katsayısının, varsayımların göz önüne alarak kullanan araştırmacıların yaptıkları çalışmalar üzerinden bu araştırma tekrar edilebilir.