



## A study of developing a Thurstone type attitude scale \*

Recep BİNDAK\*\*

Cahit PESEN\*\*\*

**ABSTRACT.** In this study a Thurstone type mathematics attitude scale is constructed and its reliability and validity is examined. The items of a Likert type math attitude scale which was developed by Aşkar (1986) are used as item pool. Forty-eight of expert group have been used as judge in this study. The ambiguous coefficients and scale values for the items are determined by the equal-appearing intervals. Reliability and validity analyses were performed by applying the scales to 66 high school students. The validity of scale was examined in terms of concurrent validity. Thurstone type and Likert type scales were compared in terms of some psychometric properties.

**Keywords:** Scaling with judgement-based, Thurstone type scale, equal-appearing intervals method, math attitude scale.

---

\*This study was presented at 2<sup>nd</sup> National Congress of Measurement and Evaluation in Education and Psychology

\*\* Assist. Prof. Dr., Gaziantep University Gaziantep, Turkey. E-mail: bindak@gantep.edu.tr

\*\*\* Assoc. Prof. Dr., Siirt University, Faculty of Education, Siirt, Turkey.

E-mail: cpesen@siirt.edu.tr

## SUMMARY

**Purpose and Significance:** The experimental methods used in scaling can be divided into two approaches. The experimental methods in the first group are the methods known as “judgment approaches”. The ones in the second group are the methods that are based on subjects’ responses, which are labeled as “response approaches”. Judgment approach is based on scaling the available stimuli according to observers’ or experts’ judgments in a dimension determined beforehand. In experimental methods, N numbers of observers (judges) each are asked to determine the stimulation degree of each of the K number of stimulants. The observers’ duty is to state the size of each stimulant with regard to scaling according to other stimulants. The average value of observer judgments for a stimulant is regarded as the scale value. Typical examples of this approach can be found in scaling that is performed through Thurstone methods (Güler&Anıl, 2009).

There are three methods of Thurstone type scaling. These paired-comparison method, successive intervals method and equal-appearing intervals method. There is also need to the judge decisions of each of the three methods. The equal-appearing intervals method was discussed in this study.

The first step in developing a Thurstone scale using method of equal appearing intervals, is to gather items about the subject to be studied. For scale value of each items it is necessary to gather rate from judges. The judges are instructed to place each items into one of a number of rating 11 intervals according to how favorable / unfavorable an evaluation it expresses. The methods assumes that the judges sort the items into what appear to them to be equal intervals. Judges are told not to express their own views about the attitude object but to judge favorableness or unfavorableness expressed by the item. Scale values for the items are easily determined. The scale value of an item is the median or mean of the scores assigned to the item by all the judges. There are two criteria for eliminating inadequate items; one is to detect ambiguous items; the second criterion is intended eliminate irrelevant items. The interquartile range is the difference between the third and first quartile ( $Q_3-Q_1$ ) is used a measure of ambiguity for each item. Highly ambiguous items would be distributed by the judges across a wide range of intervals on the evaluative continuum. Therefore, items that have high ambiguous coefficients should be eliminated because their judged favorableness varies considerably with different judges. In the final scale form items can be arranged in random order and presented to respondents whose attitudes are to be measured. They should be asked to either agree or disagree with each items. A respondent’s attitude score is the

mean or median of the scale values of the items that she/he endorses. In the process of scale construction the Thurstone method relies upon judgements of content experts to rate scale items in terms of importance. In the original technique experts were asked to sort items into 11 piles, ranked from 1 (least unfavorable) to 11 (most favorable). Each statement was then assigned a value (usually the median response of all judges to that item) which then permitted the selection of items rated the most consistently by content experts (Eagly & Chaiken, 1993; Edwards, 1994).

It is clear that there are large number of works in the field of scale development topic, specifically in attitude measurement. However, almost all of them are Likert-type scales. Thurstone-type attitude scale studies conducted in the field are negligible and this has encouraged researchers to work in this area. In this study, using Thurstone's equal-appearing intervals method, developed a math attitude scale.

The purpose of this study was to develop a Thurstone-type math attitude scale and to compare Thurstone-type , and Likert-type scales in terms of some psychometric properties. By these aspects the present study is expected to contribute to the work of scaling.

**Method:** For this study the 20 items of Likert type math attitude scale (Aşkar, 1986) was used. Primarily a data collection form was prepared for judgements. In the form there was 11 grade scale which referees point out their opinion for each attitude item. An expert grup of 48 academicians, teachers and graduate students have been used as judges in this study. The ambiguous cofficients and scale values for each item were calculated by the method of Thurstone's equal-appearing intervals. In order to gather evidence for the reliability-validity of both forms (Likert form and Thurstone form) of scale were applied to a group 9<sup>th</sup> grade students from Gaziantep-Ismetpasa High School. Students' mathematics achievement grades were collected during the implementation of this form. Data were analyzed by Microsoft Office Excel.

**Results:** According to the classification of judges the frequency table was created. From this table cumulative frequency, first, second and third quartiles were calculated by using Microsoft Office Excel. For each item the ambiguous cofficient was calculated as difference between the third and first quartile (see Table 1).

The average of median values for positive items were 9.052; and the average of median values for negative items were 2.165. The item 14 had lowest ambiguous cofficient and the item 18 had highest ambiguous cofficient. Thus the item which most agreed on by judges was item 14. In

addition, according to the judges the item 7 is located the most negative end, while item 13 and item 1 is located in the most positive end.

In this study, a Thurstone-type Mathematics attitude scale was developed and some properties were investigated. For Thurstone-type scale ambiguous coefficients were calculated as item analysis, the scale internal consistency reliability coefficient was estimated by half-splitting method. The scale reliability coefficient was calculated as 0.811. The reliability coefficient basis individual scale scores with range was 0.525. Thurstone type and Likert type scales were compared in terms of some psychometric properties. Negative and statistically significant relationship ( $n=66$ ;  $r = -0.312$ ,  $p < 0.01$ ) was found between ambiguous coefficients for Thurstone type scale and item-total correlation for Likert form. This finding support similar-scale validity. Applying Thurstone type scale and Likert form to same students, high positive and significant correlation ( $n=66$ ;  $r = 0.910$ ,  $p < 0.01$ ) were obtained. In addition, positive and significant correlation ( $n=64$ ;  $r = 0.623$ ,  $p < 0.01$ ) between Math achievement points and scale scores, was evaluated as evidence concerning the validity of the scale..

**Discussion and Conclusions:** The distributions of scores which obtained both types of scales were compared. When compared with distribution of scores obtained Likert-type scale, Distributions of scores obtained from Thurstone-type scale were more left skewed and more kurtosis. In terms of compared reliability coefficients Thurstone type scale reliability coefficient was lower than the Likert-type scale. Although stage of developing Thurstone type scale requires more steps, this situation is reverse for applying scale. Completion Thurstone-type form takes less time consuming and at same conditions is more simple to fill.



## Thurstone Tipi Bir Tutum Ölçeği Geliştirme Çalışması\*

Recep BİNDAK\*

Cahit PESEN\*\*

**ÖZ.** Bu çalışmada Thurstone'ın eşit görünen aralıklar metodu kullanılarak bir matematik tutum ölçeği geliştirilmiş ve güvenilirlik-geçerliliğine ilişkin bazı bulgular sunulmuştur. Ölçeğin maddeleri daha önce yapılmış bir çalışmada kullanılan Likert tipi matematik tutum ölçeğinden alınmıştır. Çalışmada 48 hakem yargısından elde edilen veriler kullanılarak ölçek maddeleri için belirsizlik katsayıları ve madde ölçek puanları hesaplanmıştır. Güvenirlik – geçerlik çalışması için kullanılan veriler dokuzuncu sınıfa devam eden 66 öğrenciye ölçek uygulanması ile elde edilmiştir. Ölçeğin test-yarılama güvenilirlik katsayısı ve benzer ölçekler geçerliliği için bazı kanıtlar sunulmuştur. Ayrıca Thurstone tipi ve Likert tipi matematik tutum ölçekleri psikometrik özellikleri açısından karşılaştırılmıştır.

**Anahtar Sözcükler:** Yargıcı kararlarına dayalı ölçekleme, Thurstone tipi ölçek, eşit görünen aralıklar metodu, matematik tutum ölçeği.

\* Bu çalışma Eğitimde ve Psikolojide Ölçme Değerlendirme II.Ulusal Kongresi'nde sözlü bildiri olarak sunulmuştur.

\* Yrd. Doç. Dr., Gaziantep Üniversitesi, Gaziantep, Türkiye. E-posta: bindak@gantep.edu.tr

\*\* Doç. Dr., Siirt Üniversitesi Eğitim Fakültesi, Siirt, Türkiye. E-posta: cpesen@siirt.edu.tr

## GİRİŞ

Tutum ölçümü konusunda yapılan girişimlerin çoğu, bireylerin bir dizi ifadeye verdikleri tepkilere dayalı olarak tutum hakkında çıkarsamalarda bulunmaya olanak tanıyan yöntemleri kapsar. Bu yöntemler ölçekleme teknikleri olarak tanımlanmakta ve ölçek olarak ifade edilen araçları içine almaktadır. Bu nedenle bireylerin bir dizi cümle ya da sıfat listesine gerçek duyguları doğrultusunda tepkide bulunmalarının sağlanması, *tutumların ölçülmesi* olarak anlaşılır. Bu sıfat ya da cümle listelerine ölçek adı verilir. Bu ölçeklerden ikisi Thurstone Ölçekleri ve Likert Tipi Ölçekleri'dir. Her iki ölçek türünde de tutum boyutunun orta noktasından her iki yöne uzaklık o tutumun güçlülük derecesini verir. Likert ölçeklerinde, tepki kategorilerinin bir boyut üstünde yerleştirilmesi, onları diğer ölçeklerden ayırır. Likert ölçeklerindeki tepki kategorileri, yalnızca tutum boyutunun iki ucuna yakın olacak şekilde yazılır. Thurstone ölçeklerinde tepki kategorileri, tutum boyutu üzerindeki tüm noktaları temsil edecek şekilde, ölçek üzerinde eşit görünen aralıklarla dizilip yazılmışlardır. Buna karşılık, olumlu cümlelerin tutum boyutu üzerinde belirli bir nokta etrafında toplanacağı sayılıdır (Anderson, 1991).

Psikolojik özelliklerin ölçülmesinde kullanılan ölçeklerin geliştirilmesi aşamasında başvurulan yöntemler temelde iki kaynağa dayanır. Bunlar yargıcı kararları ve denek tepkileridir (Kan, 2007). Yargı yaklaşımının esası, eldeki uyarıcıları, gözlemci yargılarına dayanarak belirlenmiş bir boyutta ölçeklemektir. Deneysel yöntemlerde, N tane gözlemcinin her birinden K tane uyarıcının her birinin uyarıcılık derecesini belli bir yöntemle belirtmesi istenir. Herhangi bir uyarıcı için gözlemci yargılarının ortalama değeri, onun ölçek değeri olarak kabul edilir. Thurstone yöntemleriyle ölçeklemede bu yaklaşımın tipik örneklerini görmek mümkündür (Anıl ve Güler, 2006).

Louis L. Thurstone (1887-1955), sosyal tutumların ölçülebileceğini ilk defa öne süren kişidir. İlk araştırmalarında “çiftli karşılaştırma” adı verilen yöntemi kullanmış, daha sonra “eşit görünen aralıklar tekniğini” geliştirmiştir (Kağıtçıbaşı, 1999:135). Eşit görünen aralıklar tekniği nesnelere sıralı bir değerlendirmesini vermesi yanında ölçek üzerindeki herhangi iki nesne arasındaki uzaklığa ilişkin yargılara varmaya da olanak vermektedir. Bu nedenle eşit görünen aralıklar tekniği tutum alanında bir çeşit ölçek geliştirme yaklaşımı sunmaktadır. Thurstone, bu teknik ile, Likert tipi ölçeklerden farklı olarak, tutumları ölçülecek bireylerin yargılarına başvurmadan önce ölçeği oluşturan maddelerin “ölçek değerlerini” bulma yöntemini getirmiştir (Özgüven, 1999:359).

Thurstone tipi ölçekleme denildiği zaman birkaç tane yöntem akla gelir. Bunlar; ikili karşılaştırma yöntemi, ardışık aralıklar yöntemi ve eşit görünen aralıklar yöntemidir. Her üç yöntemde de yargıcı kararlarına gerek vardır

(Roberts et al., 1999). Bu çalışmada, Thurstone tipi ölçekleme kapsamında eşit görünen aralıklar yöntemi kullanılmıştır.

Literatür incelendiğinde, Thurstone ve Likert tipi ölçeklerin karşılaştırılması üzerinde duran pek çok çalışmaya rastlanmıştır. Bu çalışmaların ortak sonuçları; 1) Thurstone tipi ölçek geliştirmenin Likert tipi ölçek geliştirmeye göre daha fazla çaba gerektirdiğini, 2) Maddelerin ölçek değerini belirlemede hakem yargısı kullanıldığından hakemlerin kendi tutumlarını da yansıtmış olma olasılığı bulunduğunu, 3) Madde puan (ölçek değerleri) varyanslarının toplam ölçek puanına yansımadığını, 4) Güvenirlik katsayıları bakımından karşılaştırıldıklarında Thurstone tipi ölçeklerde, Likert tipi ölçeklere göre daha düşük değer elde edildiğini ortaya koymuştur (Edwards ve Kenney, 1946; Kağıtçıbaşı, 1999; Özgüven, 1999:361; Roberts et al., 1999). Thurstone tipi ölçekler için bir diğer dezavantaj olarak, ölçek konularının kültürler arasında geçerliliğinin olmadığı şeklinde bir eleştiri yapılmışsa da bu durumun bütün ölçekler için geçerli olduğu unutulmamalıdır (Tavşancıl, 2002:137)

Thurstone tipi ölçek geliştirme aşamalarını şu şekilde sıralamak mümkündür (Anderson, 1999; Eagly & Chaiken, 1993; Edwards, 1994; Kağıtçıbaşı, 1999; Özgüven, 1999; Tavşancıl, 2002). Tutum konusuna ilişkin çok sayıda ifade hazırlanır, tutum maddesinin ölçek sıralamasındaki yerini saptamak üzere bir uzmanlar (yargıcılar) grubuna sunulur. Uzmanlardan her bir ifadeyi 11 gruptan birine atamaları istenir, uzmanlarca birbirinden tutarsız yerlere yerleştirilen ifadeler iptal edilir, uzmanların bir madde için verdikleri derecelerin medyan değeri veya aritmetik ortalaması bu maddenin ölçek değeri olarak kabul edilir.

Yargısına başvurulacak uzman sayısının ölçek değerleri üzerinde çok az etkisinin bulunduğu rapor edilmiştir. Thurstone ve Chave (1929) tarafından yapılan çalışmada 129 maddelik bir ölçek için 72 uzman yargısından elde edilen ölçek değerleri ile 300 uzman yargısından elde edilen ölçek değerleri arasında .95 korelasyon bulunmuştur (Akt. Edwards A.L. (1994:94). Ancak ölçek geliştirme sürecinde yargısına başvuru alanların sayısının 15'ten az olmaması önerilmektedir (Eagly & Chaiken, 1993:43).

Thurstone tipi ölçeklere madde seçmede hakem yargıları arasındaki tutarlılığın bir ölçüsü olan belirsizlik katsayısı kullanılmaktadır. Belirsizlik katsayısı, belirsiz ve konu dışı maddeleri elemeye kullanılacak bir ölçüttür (Tavşancıl, 2002;134).

Thurstone tipi ölçeklemede belirsizlik katsayısı hesaplanırken öncelikle yargıcı görüşlerinden her bir madde için birikimli frekans dağılımı oluşturulur. Birikimli dağılımda, %25'inin altında kalan ve %75'inin üstünde kalan gözlemlerin farkı, maddenin bir istatistiksel belirsizlik ölçütünü ifade etmek için kullanılır. Bu farka çeyrekler açıklığı ( $Q=inter-$

*quartile range*) değeri denir. Küçük belirsizlik katsayısı [Q] değeri ortadaki %50'lik hakem yargılarının küçük bir aralığa dağıldığını gösterir, dolayısıyla bu da hakem yargıları arasında bir tutarlılığı işaret eder (Edwards ve Kenney, 1946; Guffey et al., 2007). Dikkat edilirse belirsizlik katsayısı birikimli frekans tablosunda üçüncü ile birinci çeyreklikler arası farktan başka bir şey değildir.

Hesaplanan Q değerleri maddelerin göreceli olarak ayırıcılık indeksleri olarak okunabilir. Yani düşük Q değeri maddenin ayırıcılık gücünün yüksekliğini göstermektedir. Bir madde için büyük Q değeri yargıcı kararları arasında tutarsızlık bulunduğu işaret eder, böyle bir durumda maddenin gözden geçirilmesi veya ölçekten atılması gerekir (Edwards, 1994:89). Ancak seçilecek maddeler için Q değerinin üst sınırı konusunda çeşitli öneriler bulunmaktadır. En yaygın görüş Q değerinin 2.5'ten büyük olmaması (Thurstone,1959'dan akt: Rainey, 2002) şeklindedir ancak Q için kabul edilebilir üst sınır olarak 4 değerinin benimsendiği çalışmalara da (Guffey et al., 2007; Patel ve Chauhan, 2009; Rainey, 2002) rastlamak mümkündür.

Thurstone ölçekleri için güvenilirlik kestirme alfa, yarıya bölme, madde-toplam korelasyonu yöntemleri maddelerinin ölçek değerleri farklı olması nedeniyle uygun değildir. Ölçek derecelerinin 2'li olması, bu ölçeklerde KR-20 veya KR-21 güvenilirlik analizlerinin uygulanabileceği anlamına da gelmez (Şencan, 2005:87). Thurstone ölçekleri için güvenilirliğin kestirilmesinde ölçeğin iki paralel formu arasında hesaplanacak korelasyon katsayısının kullanımı veya test-tekrar test yönteminin uygulanması önerilmektedir (Edwards ve Kelley, 1946; Edwards 1994:94; Şencan, 2005:87). Anderson (1991), Thurstone ölçekleri için geleneksel iç tutarlılık göstergeleri yerine her birey için ayrı ayrı hesaplanabilecek bir güvenilirlik ölçütü önermektedir. Bu ölçüt bireyin onayladığı cümlelerin ölçek değerlerinin ranjına bakılarak bireylerin cevaplarındaki tutarlılık derecesinin kestirilmesi şeklindedir.

Matematiğe yönelik olumlu tutum geliştirmek, matematik eğitiminin genel amaçlarından birisidir (MEB, 2009). Bu nedenle öğrencilerin matematik dersine ve/veya matematiğe yönelik tutumlarının ölçülmesi önemli bir konu olarak karşımıza çıkmaktadır. Nitekim öğrencilerin matematiğe yönelik tutumlarının ölçüldüğü çok sayıda çalışmaya rastlamak mümkündür. Özel olarak matematik tutum ölçeği geliştirmek amacıyla yapılmış çalışmalar (Aşkar, 1986; Duatepe ve Çilesiz, 1999; Erol, 1989; Nazlıçipek ve Erkin, 2002) bulunmaktadır. Söz konusu bu ölçeklerin tümü Likert tipi ölçeklerdir.

Yurtiçinde eşit görünen aralıklar metodu ile geliştirilen Thurstone tipi ölçeklerin iş yaşamında kullanıldığı görülmektedir (Kaya ve Kaya 2007; Özdevecioğlu, 2002). Ancak tutum ölçme konusunda yurtiçinde yapılmış



herhangi bir araştırmaya rastlanmamıştır. Yurtdışında ise çeşitli konulara yönelik tutum ölçme aracı geliştirilmesinde Thurstone'un eşit görünen aralıklar yöntemi kullanılmıştır. Bunlardan bazıları müzik dersine (Raney, 2002), feminizm ve kadın hareketlerine (Fassinger, 1994), kan bağışına (Breckler & Wiggins, 1989) ve zenci öğrencilere (Grafton, 1964) yönelik tutum ölçekleridir.

Ölçme hem fiziksel bilimlerde hem de sosyal bilimlerde önemli araçlardan biridir. Sosyal bilimlerde özellikle psikolojik özelliklerin ölçülmesinde kullanılacak ölçeklerin geliştirilmesi aşamasında kullanılacak birçok yöntem geliştirilmiştir. Bu yöntemler, biri yargıcı kararları, diğeri denek tepkileri olmak üzere temelde iki kaynağa dayalıdır. Her iki yöntemde, sosyal bilimler alanındaki araştırmalarda ölçme aracı geliştirmek için başvurulan yöntemlerdir (Kan, 2008). Tutumları ölçen ölçme araçlarının geliştirilmesi, tutumların bilimsel yöntemlerle incelenerek daha tutarlı bir tutumlar kuramının oluşmasına yol açmakta ve kişiler hakkındaki çeşitli kararlarda tutum puanlarının kullanılmasına da olanak vermektedir (Tavşancıl, 2002). Yurtiçinde ve yurtdışında ölçek geliştirme ve özel olarak tutum ölçme alanında yapılmış çalışmalara bakıldığında hemen hemen tamamında denek tepkilerine dayalı yöntemlerin kullanıldığı görülmektedir. Likert tipi ölçek geliştirme çalışmaları buna en tipik örnektir. Buna karşılık yargıcı kararlarına dayalı yöntem kullanılan çalışmalar çok nadirdir. Yargıcı kararlarına dayalı olan Thurston'un eşit görünen aralıklar yöntemi kullanılarak bir tutum ölçeği geliştirme çalışmasının alana katkı getirmesi bakımından önemli olduğu düşünülmüştür. Bu nedenle bu çalışmada Thurstone tipi bir matematik tutum ölçeğinin geliştirilme aşamaları ele alınmış ve Thurstone tipi ölçek ile Likert tipi ölçek çeşitli psikometrik özellikleri bakımından karşılaştırılmıştır. Bu yönüyle çalışmanın tutum ölçme ve/veya ölçekleme alanında çalışan araştırmacılara katkı sağlaması beklenmektedir.

Bu çalışmanın temel amacı Thurstone tipi bir matematik tutum ölçeği geliştirmek, Thurstone tipi ve Likert tipi matematik tutum ölçeklerini psikometrik özellikleri açısından karşılaştırmaktır.

## **YÖNTEM**

Bu çalışma için bir madde havuzu oluşturulmamış, bunun yerine Aşkar (1986) tarafından geliştirilen 20 maddelik Likert tipi matematik tutum ölçeğinin maddeleri kullanılmıştır. Ölçeğin maddelerinden 10 tanesi olumsuz tutum cümlesi biçimindedir. Ölçek, geliştirilme aşamasında 240 öğrenciye uygulanmış, bu uygulamadan elde edilen veriler kullanılarak geçerlik ve güvenirlik çalışmaları yapılmıştır. Ölçeği oluşturan maddelerin

faktör yüklerinin 0.63 ile 0.86 arasında değiştiği, ölçeğin tek faktörlü olduğu ve Cronbach alfa iç tutarlılık katsayısının 0.96 olduğu rapor edilmiştir (Aşkar, 1986). Bu çalışmada ölçeğin Thurstone tipi formunun geliştirilmesi sürecinde öncelikle hakem yargılarını elde etmek için bir veri toplama formu hazırlanmıştır. Formda her bir tutum maddesinin karşısında hakemlerin yargılarını belirtmeleri için 11 dereceli bir ölçek yer almıştır. Hakemlerden söz konusu tutum maddelerini ölçek üzerinde uygun konuma atamaları, her bir ifade için yargılarını ve tutumun derecesini belirlerken kendi tutumlarını değil genel olarak matematik tutumuna ilişkin yargıları belirtmeleri istenmiştir.

1	2	3	4	5	6	7	8	9	10	11
Olumsuz			Nötr					Olumlu		

### Çalışma Grubu

Hakem olarak yargısına başvuru alanlar matematik eğitimi veya eğitim bilimleri alanında çalışan akademisyenler, sınıf öğretmenleri, matematik öğretmenleri ve Eğitim Bilimleri alanında yüksek lisans/doktora yapmakta olan öğrencilerden oluşmuştur. Bu şekilde yargısına başvuru alan hakem sayısı 48'dir. Thurstone tipi ölçek için maddelerin belirsizlik katsayısı ile madde ölçek puanları hesaplanmıştır. Güvenilirlik-geçerliliğe ilişkin kanıt toplamak amacıyla 2009-2010 bahar yarıyılında Gaziantep ilindeki İsmetpaşa Lisesi'ne devam eden bir grup dokuzuncu sınıf öğrencisine ölçeğin hem Likert formu ve hem de ikili-Thurstone formu uygulanmıştır. Öğrencilerin karne notları bu formun uygulanması sırasında toplanmıştır. Veriler Microsoft Excel programı ile analiz edilmiştir.

### Madde Analizi

Thurstone tipi ölçekte yer alacak maddelerin uygunluğu için madde belirsizlik katsayıları hesaplanmıştır. Bilindiği gibi belirsizlik katsayısı bir tutum maddesi üzerinde ne ölçüde uzlaşma olduğunu gösteren ve sayısal olarak ölçülebilen bir değerdir.

Hakemlerin sınıflamasına göre 20 tane maddenin aldığı ölçek konumu için Ek-1'deki frekans cetveli oluşturulmuştur. Microsoft Excel programı kullanılarak frekans cetvelinden yığılmalı frekanslar ile birinci, ikinci ve üçüncü çeyreklik değerleri hesaplanmıştır. Her bir madde için belirsizlik katsayısı olarak üçüncü çeyreklik ile birinci çeyreklik arasındaki fark hesaplanmıştır. Çeyreklik için

$$Q = A_s + \frac{(pN - f_a)}{f_s}$$

formülü kullanılmıştır. Burada  $A_s$ ; kümenin gerçek sınır,  $p$ ; oran (0.25, 0.50 veya 0.75),  $N$  ; toplam veri sayısı,  $f_a$ ; birikimli frekans sütununda  $pN$ 'nin kapsandığı değer altındaki toplam frekans,  $f_s$  ; çeyrekliklerin bulunduğu değer frekansıdır. Bir maddenin medyanı,  $p=0.5$  alınarak hesaplanan ikinci çeyreklik değeridir.  $p=0.25$  için  $Q_1$  ve  $p=0.75$  için  $Q_3$  hesaplanarak belirsizlik katsayısı olarak  $Q=Q_3-Q_1$  değerleri Tablo-1'deki gibi bulunmuştur.

**Tablo 1.** Ölçek maddelerinin belirsizlik katsayıları ve maddelerin medyan değerleri

Madde No	Q1	Q3	Belirsizlik katsayısı (Q)	Madde-Ölçek değeri
1	7.928	10.642	2.714	9.500
2n	1.250	3.166	1.917	2.300
3n	1.357	3.428	2.071	2.571
4	7.167	10.300	3.133	9.100
5	6.375	10.192	3.817	9.000
6n	1.423	3.625	2.202	2.187
7n	1.000	2.750	1.75	1.500
8	7.785	10.357	2.571	8.750
9n	1.300	2.966	1.667	2.143
10n	1.766	4.166	2.400	2.611
11	6.833	9.786	2.952	8.667
12n	1.166	3.125	1.958	1.961
13	7.071	10.833	3.762	9.500
14	8.562	10.100	1.537	9.300
15n	1.206	3.346	2.140	2.375
16n	1.300	3.591	2.291	2.100
17	8.083	9.928	1.845	9.038
18	6.722	10.577	3.855	8.500
19n	1.167	2.833	1.667	1.900
20	7.166	10.269	3.103	9.166

n: Negatif madde

Ölçekte yer alan olumlu maddelerin medyan değerleri ortalaması 9.052, olumsuz maddeler için medyan değerleri ortalaması 2.165 olarak elde edilmiştir. Tablo1'de görüldüğü gibi en düşük belirsizlik katsayısı 14.madde, en yüksek belirsizlik katsayısı ise 18.madde için hesaplanmıştır. Belirsizlik katsayısı sifra yaklaştıkça uzlaşmanın arttığı düşünüldüğünde hakemler tarafından üzerinde en çok uzlaşılan maddenin 14 ve en az uzlaşılan

maddenin de 18 olduğu anlaşılmaktadır. Ayrıca, hakem yargılarına göre 7.madde en olumsuz uçta yer alırken 1 ve 13.maddeler en olumlu uçta yer almaktadır. Tüm maddeler için belirsizlik katsayısının 4'ün altında olduğu ve kullanılan maddelerin daha önce geliştirilmiş Likert tipi ölçekten alındığı göz önüne alınarak madde azaltma yoluna gidilmemiştir.

Ölçeğin Likert formu ile Thurstone formu, 66 kişilik bir grup dokuzuncu sınıf öğrencisine uygulanmıştır. Her iki tip ölçeğe ilişkin hesaplanan bazı betimsel istatistikler Tablo-2'de özetlenmiştir.

**Tablo 2. Ölçek için bazı istatistikler**

	Likert tipi ölçek	Thurstone tipi ölçek
Ortalama	3.305	5.994
Standart sapma	0.885	2.487
En küçük puan,	1.5	1.85
En büyük puan	4.95	9.12
Çarpıklık	0.044	-0.100
Basıklık	-0.824	-1.561
Test yarılama güvenilirlik katsayısı	0.934	0.811
Öğrencilerin ölçekten aldıkları tutum puanları ile matematik dersi karne notları arasındaki korelasyon n=64*	0.657 ; p<0.01	0.623; p<0.01
Öğrencilerin Ölçeğin iki formundan aldıkları puanlar arası korelasyon n=66		0.910

\*Karne notu tespit edilemeyen 2 kişi var

### Güvenirlilik Çalışmaları

Thurstone tipi matematik tutum ölçeğinin güvenilirlik katsayısı, ölçek, tek/çift numaralı maddeler şeklinde iki yarıya bölünerek tahmin edilmiştir. Sonuçta ölçeğin test yarılama güvenilirlik katsayısı Spearman-Brown düzeltmesi ile 0.811 olarak bulunmuştur. Anderson (1991) tarafından önerilen güvenilirlik ölçütü ise 0.525 olarak bulunmuştur. Bu katsayı bulunurken bireylerin ölçek puanları ranjı ortalaması (5.30) ile en küçük teorik ranjin (1.50) farkı, mümkün olan teorik en büyük ranj (8.00) değerine oranlanmış ve sonuç 1'den çıkarılmıştır. Literatürde birey ölçek puanları ranjına dayalı güvenilirlik katsayısına dair herhangi bir bulguya rastlanmadığından test yarılama yöntemi ile bulunan güvenilirlik katsayısı ile birey ölçek puanları ranjına dayalı güvenilirlik katsayısı arasındaki bu denli büyük farkın nedeni açıklanamamıştır.

**Tablo 3.** Thurstone tipi ölçek için *Q* değerleri ile likert tipi ölçek için madde-toplam puan korelasyonları

	Q değeri	Madde-toplam kore.
1. Matematik sevdiğim bir derstir	2.714	0.747
2. Matematik dersine girerken büyük bir sıkıntı duyarım	1.917	0.589
3. Matematik dersi olmasa öğrencilik hayatı daha zevkli olur	2.071	0.582
4. Arkadaşlarımla matematik tartışmaktan zevk duyarım	3.133	0.603
5. Matematığe ayrılan ders saatlerinin fazla olmasını dilerim	3.817	0.549
6. Matematik dersi çalışırken canım sıkılır	2.202	0.703
7. Matematik dersi benim için bir angaryadır	1.750	0.555
8. Matematikten hoşlanırım	2.571	0.687
9. Matematik dersinde zaman geçmek bilmez	1.667	0.667
10. Matematik dersi sınavından çekinirim	2.400	0.426
11. Matematik benim için ilgi çekicidir	2.952	0.742
12. Matematik bütün dersler içinde en çok korktuğum derstir	1.958	0.609
13. Yıllarca matematik okusam bıkmam	3.762	0.752
14. Diğer derslere göre matematiği daha çok severek çalışırım	1.537	0.686
15. Matematik beni huzursuz eder	2.140	0.708
16. Matematik beni ürkütür	2.291	0.635
17. Matematik dersi eğlenceli bir derstir	1.845	0.693
18. Günlük hayatta matematik, çok işimize yarar	3.855	0.156
19. Derslerin içinde en sevimsizi matematiktir	1.667	0.641
20. Çalışma zamanımın çoğunu matematiğe ayırmak isterim	3.103	0.627

Tabloda görüldüğü gibi belirsizlik katsayısı 4'ün üzerinde olan hiçbir madde bulunmamaktadır. Ayrıca ölçeğin Likert formu için 18. madde dışındaki tüm maddeler için madde toplam korelasyon değerleri 0.01 düzeyinde anlamlı bulunmuştur.

### Geçerlik Çalışmaları

Thurstone tipi ölçeğin uygulanması sonucu elde edilen matematik tutum puanları ile karne notları arasındaki korelasyona bakılmıştır. Tutum puanları ile matematik karne notları arasında istatistiksel olarak anlamlı bir korelasyon bulunmuştur ( $n=64$ ,  $r=0.623$ ;  $p<0.01$ ). Bu durum ölçeğin geçerliğine ilişkin bir kanıt olarak değerlendirilmiştir.

Matematik tutum ölçeğinin Thurstone formu ve Likert formundan elde edilen matematik tutum puanları arasındaki korelasyona bakılmış, her iki formdan elde edilen puanlar arasında yüksek derecede pozitif ve istatistiksel olarak anlamlı ( $n=66$ ;  $r=0.910$ ;  $p<0.01$ ) bir korelasyon elde edilmiştir. Ancak her iki ölçeğin madde ifadelerinin aynı olduğu dikkate alınarak bu korelasyonda şartıcı değişken olabileceği varsayılarak ölçekteki maddeler teker teker karşılaştırılmıştır. Bu amaçla ölçeğin Likert formundan elde

edilen madde-toplam korelasyon değerleri ile Thurstone formundan elde edilen belirsizlik katsayıları karşılaştırılmıştır. Sonuçta negatif ve anlamlı korelasyon elde edilmiştir ( $n=66$ ;  $r=-0.312$ ;  $p<0.01$ ). Thurstone tipi ölçeklerde belirsizlik katsayısının 0'a yakın ve Likert tipi ölçeklerde madde-toplam puan korelasyonunun ise 1'e yakın olması ölçek güvenilirliğinin yüksek olması anlamına geldiğine göre her iki ölçek tipi arasında hesaplanan korelasyon katsayısının negatif ve anlamlı olması benzer ölçekler geçerliğinin sağlandığı şeklinde değerlendirilebilir.

### Ölçek Puanlarının Değerlendirilmesi

Thurstone tipi ölçek, tutumları ölçülecek gruba uygulanırken ölçek puanları belirtilmez. Cevaplayıcıdan; maddelere, “Evet/Hayır” veya “benim için doğru /benim için doğru değil” veyahut “katılıyorum/katılmıyorum” biçiminde tepki vermeleri istenir. Sonuçta bireyin “Evet” olarak belirttiği maddelerin değerleri toplanır ve sonuç “Evet” olarak belirtilen madde sayısına bölünür. Elde edilen değer, kişinin tutum puanıdır.

### TARTIŞMA VE SONUÇ

Bu çalışmada Aşkar (1986) tarafından geliştirilmiş olan Likert tipi matematik tutum ölçeğinin maddeleri kullanılarak Thurstone tipi bir matematik tutum ölçeği geliştirilmiştir. Thurstone tipi ölçek için madde analizi olarak belirsizlik katsayıları hesaplanmış, ölçeğin iç tutarlılık güvenilirlik katsayısı testi yarılama yöntemi ile tahmin edilmiştir. Ölçeğin güvenilirlik katsayısı 0.811 olarak hesaplanmıştır. Birey ölçek puanları ranjı ile bulunan güvenilirlik katsayısı ise 0.525 bulunmuştur. Belirsizlik katsayıları ile ölçeğin Likert formundan elde edilen madde-toplam korelasyon değerleri arasında bulunan negatif ve istatistiksel olarak anlamlı ilişki ( $r=-0.312$ ;  $p<0.01$ ) ölçeğin benzer ölçekler geçerliğini desteklemektedir. Ölçeğin güvenilirlik ve geçerliğine ilişkin başka kanıtlar için ölçeğin Thurstone formu ile Likert formu aynı bireylere uygulanmış ve bu iki uygulama arasında yüksek düzeyde pozitif ve istatistiksel olarak anlamlı bir korelasyon ( $r=0.910$ ;  $p<0.01$ ) bulunmuştur. Ayrıca her bir öğrencinin Thurstone tipi ölçek ile elde edilen tutum puan ortalaması puan ortalaması ile matematik dersi karne notları arasında bulunan pozitif ve anlamlı ( $r=0.623$ ;  $p<0.01$ ) ilişki de ölçeğin geçerliğine ilişkin bir kanıt olarak değerlendirilmiştir.

Öğrencilerin her iki tip ölçekten aldıkları puanların dağılımı karşılaştırıldığında Thurstone tipi ölçekten alınan puanların Likert tipi ölçekten alınan puanlara göre daha çok sola çarpık olduğu ve Thurstone tipi ölçekten alınan puanların dağılımının Likert tipi ölçekten alınan puanların dağılımına göre daha basık olduğu görülmüştür.

Güvenirlilik katsayıları bakımından karşılaştırıldığında Thurston tipi ölçeğin güvenirlilik katsayısı Likert tipi ölçeğe göre daha düşük bulunmuştur. Ayrıca öğrencilerin karne notları ile tutum ölçeklerinden aldıkları toplam puanların ilişkisine bakıldığında yine Thurstone tipi ölçek için hesaplanan korelasyon katsayısı Likert tipi ölçek için hesaplanan korelasyon katsayısına göre daha düşük bulunmuştur. Bu durum literatürde yer alan Thurstone ölçeklerinin güvenirliliklerinin aynı şartlarda Likert tipi ölçeklere göre daha düşük olduğu şeklindeki bulguları (Edwards ve Kenney, 1946; Kağıtçıbaşı, 1999; Özgüven, 1999:361; Robers et al., 1999) desteklemektedir.

Her ne kadar Thurstone tipi ölçeğin geliştirilmesi ve puanlanması aşamaları Likert tipi ölçeğe göre daha çok adım gerektiriyorsa da ölçeğin uygulaması için bu durum tersinedir. Thurstone tipi ölçeğin doldurulması daha basittir ve aynı şartlarda daha az zaman almaktadır. Nitekim bir ölçeğin 2'li ve 5'li formlarının karşılaştırılması ile ilgili yapılan bir araştırmada (Atılğan, ve Saçkes, 2003) 13-14 yaş aralığında iki kategorili ölçeklemenin daha kullanışlı olduğu, öğrencilerin tereddütsüz işaretleme yaptıkları bulgusuna ulaşılmıştır. Bu nedenle özellikle ilköğretim-lise öğrencileri için Thurstone tipi ölçeğin daha uygun olacağı açıktır.

#### KAYNAKLAR

- Anderson, L.W. (1991). *Tutumların ölçülmesi* (Çeviren: N.Çıkrıkçı). *Attitudes and their measurement* In: Keeves P.J. (Ed.) *Educational Research, Methodology and Measurement: An International Handbook*. Pergamon Press, Canada  
[<http://dergiler.ankara.edu.tr/dergiler/40/506/6149.pdf>]
- Anıl, D. ve Güler, N. (2006). İkili karşılaştırma yöntemi ile ölçekleme çalışmasına bir örnek. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 30, 30–36.
- Aşkar, P. (1986). Matematik dersine yönelik tutumu ölçen Likert tipi bir ölçeğin geliştirilmesi. *Eğitim ve Bilim*, 62, 5-13.
- Atılğan, H. ve Saçkes, M. (2003). *Ölçeklerin ikili ve çok kategorili puanlanmasının psikometrik özelliklerinin karşılaştırılması*. VII. Ulusal Psikolojik Danışma ve Rehberlik Kongresi, İnönü Üniversitesi, Malatya.
- Breckler, S. J. & Wiggins, E. C. (1989). Scale of the measurement of attitude toward blood donation. *Transfusion*, 29, 401-404.
- Duatepe, A. ve Çilesiz Ş., 1999. Matematik tutum ölçeği geliştirilmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 16-17, 45-52.
- Eagly, A.H. & Chaiken, S. (1993). *The Psychology of Attitudes*. USA: Harcourt Brace Company.
- Edwards A.L. (1994). *Techniques of Attitude Scale Construction*. New York: Irvington Pub. Inc.
- Edwards, A.L. & Kenney, K.C. (1946). A comparison of the Thurstone and Likert techniques of attitude scale construction. *Journal of Applied Psychology*, 30, 72-83.

- Erol, E. (1989). *Prevalence, and correlates of math anxiety in Turkish high school students*. Yüksek Lisans Tezi, Boğaziçi Üniversitesi Fen Bilimleri Enstitüsü.
- Fassinger, R. (1994). Development and testing of the attitudes toward feminism and the women's movement scale. *Psychology of Women Quarterly*, 18, 389-402.
- Grafton, T. H. (1964). An attitude scale on accepting negro students. *Social Forces*, 43 (1), 38-41.
- Guffey, J.E., Larson, J.G., Zimmerman, L. & Shook, B.(2007). The Development of a Thurstone Scale for Identifying Desirable Police Officer Traits. *J.Police Crim Psych.* 22, 1-9.
- Güler, N., Anıl, D. (2009). Scaling through pair-wise comparison method in required characteristics of students applying for post graduate programs. *International Journal of Human Sciences* [Online]. 6 (1), 627-639. <http://www.insanbilimleri.com/en>
- Kağıtçıbaşı, Ç. (1999). *İnsan ve İnsanlar (11.Baskı)*. İstanbul: Evrim Yayınları.
- Kan, A. (2008). Yargıcı kararlarına dayalı ölçekleme yöntemlerinin karşılaştırılması üzerine ampirik bir çalışma. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 35, 186-194.
- Kaya A ve Kaya Y. (2007). Küçük ve orta boy işletme yöneticilerinin stres kaynaklarını tespit etmeye yönelik Kayseri ilinde bir araştırma. *Erciyes Üniversitesi SBE Dergisi*, 23, 41-62.
- Nazlıççek, N. ve Erktin, E. (2002). *İlköğretim matematik öğretmenleri için kısaltılmış matematik tutum ölçeği*. V. Ulusal Fen Bilimleri ve Matematik Eğitimi Kongresi, ODTÜ; Ankara. (4.2.2003), [www.fedu.metu.edu.tr/ufbmek/5b\\_kitabi](http://www.fedu.metu.edu.tr/ufbmek/5b_kitabi)
- Özdevecioğlu, M. (2002). Kamu ve özel sektör yöneticileri arasındaki Davranışsal çalışma koşulları ve kişilik farklılıklarının belirlenmesine yönelik bir araştırma. *Erciyes Üniversitesi İİBF Dergisi*, 19, 115-134.
- Özgüven, İ.E. (1999). *Psikolojik Testler*. Ankara: Pdrem Yayınları.
- Patel, M. ve Chauhan, N.B. (2009). Development of scale to measure agri-business anxiety of youth. *Agric.Sci.Digest*, 29 (3), 21-220.
- Rainey, V.J. (2002). *The Development of the "Rainey Musical Attitude Scale," Using the Thurstone Scale as a Model, to Measure Attitudes of Music Educators and Principals toward the Value of Music in the North Carolina Public School Curriculum*. Ph.D., UNC-G, 2002. 143 pp
- Roberts, J.S., Laughlin, J.E & Wedell, D.H. (1999). Validity issues in the Likert and Thurstone approaches to attitude measurement. *Educational and Psychological Measurement*, 59 (2), 211-233.
- Şencan, H. (2005). *Sosyal ve Davranışsal Ölçmelerde Güvenirlilik ve Geçerlilik*. Ankara: Seçkin Yay.
- Tavşancıl, E. (2002). *Tutumların Ölçülmesi ve Spss ile Veri Analizi*. Ankara: Nobel.
- Thurstone, L.L. (1959). *The Measurement of values*, IL: The University of Chicago Pres.
- \_\_\_\_\_ MEB (2009). *İlköğretim Matematik Dersi Öğretim Programı*. Ankara: Milli Eğitim Bakanlığı.



**EK-1. Hakem görüşlerine göre maddelerin yer aldıkları sınıflar**

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
<b>1.</b>	0	16	14	0	0	13	24	0	15	8	0	18	0	0	17	15	0	0	18	0
<b>2.</b>	0	10	9	0	1	16	10	1	14	15	0	13	1	0	8	15	0	0	15	1
<b>3.</b>	0	15	14	0	3	6	8	0	15	9	0	8	0	1	13	5	0	0	9	0
<b>4.</b>	0	4	6	1	0	8	4	0	1	6	1	6	0	0	5	11	0	0	5	0
<b>5.</b>	1	1	3	2	1	2	1	1	1	3	0	1	2	1	3	1	1	2	0	1
<b>6.</b>	3	1	0	5	8	0	0	2	1	5	9	0	5	1	1	1	0	8	1	6
<b>7.</b>	2	1	1	6	5	0	0	4	0	2	6	1	7	0	1	0	4	9	0	6
<b>8.</b>	8	0	0	4	3	2	1	14	0	0	6	0	5	8	0	0	12	5	0	6
<b>9.</b>	10	0	0	10	6	0	0	8	0	0	12	0	4	16	0	0	13	7	0	6
<b>10.</b>	10	0	1	10	13	0	0	7	0	0	7	0	6	15	0	0	14	4	0	13
<b>11.</b>	14	0	0	10	8	1	0	11	1	0	7	1	18	6	0	0	4	13	0	9

