EJSSE

Eurasian Journal of Sport Sciences and Education
Avrasya Spor Bilimleri ve Eğitim Dergisi
Vol: 6, Issue: 2, 2024
E-ISSN: 2717-8455
URL: http://www.dergipark.org.tr/ejsse

Perceived Halo Effect in Referee Evaluations in Sport: A Scale Development Study*

Yusuf Çağrı İPİN[1†], Hakan ÜNAL[2]
[1]Muğla Sıtkı Koçman University, Institute of Health Sciences, Muğla, Türkiye
[2]Muğla Sıtkı Koçman University, Faculty of Sport Sciences, Muğla, Türkiye

Original Article **DOI: 10.47778/ejsse.1521020**
Received: 23.07.2024 Accepted: 27.08.2024 Online Published: 30.09.2024

Abstract

This study was conducted to develop a scale that can measure the relationship between the Halo effect and the evaluations made by referees about athletes' performances, who should make subjective evaluations in sport competitions. The population of the study consisted of 247 referees who were on duty in various sports and in different classifications and who were registered for the year 2023. The items of the scale were produced by the researcher with the support of the literature and it was determined that the items provided content validity (CVI=0.925). After the pilot study, reliability analysis was performed and Alpha=0.970 was found. Exploratory factor analysis was applied to reveal the construct validity of the scale and Alpha=0.972 was found. The fit statistics calculated by confirmatory factor analysis were found to be compatible with the previously determined factor structure of the scale at an acceptable level. In the results of the analyses, it was determined that factor loadings were high, standard error values were low, t values were significant and the construct validity of the predetermined factor structure was confirmed. In addition, as a result of the item discrimination test, it was determined that the scale was able to make a sensitive discriminative measurement ($p<0.05$). It was found that the scale provided convergent and divergent validity (CR>AVE>0.5), ICC values related to the agreement between test-retest measurements were high, there was no difference between test-retest correlation values, and the scale made reliable measurements based on short time ($p<0.05$). As a result, it was concluded that the "Perceived Halo Effect Scale in Referee Evaluations in Sport" is a valid and reliable scale.
Keywords: Halo effect, Sport, Prejudice, Referee evaluations, Scale development

Sporda Hakem Değerlendirmelerinde Algılanan Halo Etkisi: Ölçek Geliştirme Çalışması

Öz

Bu araştırma, spor müsabakalarında sübjektif değerlendirmeler yapması gereken hakemlerin sporcu performansları hakkında yaptıkları değerlendirmeler ile Halo etkisi arasındaki ilişkiyi ölçebilecek bir ölçek geliştirmek amacıyla yapılmıştır. Çalışmanın evrenini, çeşitli spor dallarında ve farklı klasmanlarda görev alan 2023 yılı için vizeli 247 hakem oluşturmaktadır. Ölçek maddeleri literatür destekli olarak araştırmacı tarafından üretilmiş ve maddelerin kapsam geçerliliğini sağladığı saptanmıştır (KGİ=0.925). Yapılan pilot çalışma sonrası ölçeğe ilişkin güvenirlik analizi uygulanmış ve Alpha=0.970 olarak bulunmuştur. Ölçeğin yapı geçerliliğini ortaya koymak için açıklayıcı faktör analizi uygulanmış ve Alpha=0.972 olarak bulunmuştur. Doğrulayıcı faktör analizi ile hesaplanan uyum istatistiklerinin ölçeğin daha önce belirlenen faktör yapısı ile kabul edilebilir düzeyde uyumlu olduğu saptanmıştır. Analiz sonuçlarında, faktör yüklerinin yüksek, standart hata değerlerinin düşük, t değerlerinin anlamlı olduğu belirlenmiş ve önceden belirlenen faktör yapısına ilişkin yapı geçerliliği doğrulanmıştır. Ayrıca uygulanan madde ayırt edicilik testi sonucu ölçeğin ayırt edici hassas ölçüm yapabildiği saptanmıştır ($p<0.05$). Ölçeğin, yakınsak ve ayrışma geçerliliğini sağladığı görülürken (CR>AVE>0.5), test-tekrar test ölçümleri arasındaki uyuma ilişkin ICC değerlerinin yüksek bulunduğu, test-tekrar test korelasyon değerleri arasında fark olmadığı ve ölçeğin kısa zamana bağlı olarak güvenilir ölçüm yaptığı saptanmıştır ($p<0.05$). Sonuç olarak, "Sporda Hakem Değerlendirmelerinde Algılanan Halo Etkisi Ölçeği' nin geçerli ve güvenilir bir ölçek olduğu kanısına varılmıştır.
Anahtar kelimeler: Halo etkisi, Spor, Ön yargı, Hakem değerlendirmeleri, Ölçek geliştirme

* This study is derived from the first author's master's thesis titled "Perceived Halo Effect in Referee Evaluations in Sport: A Scale Development Study".

† **Corresponding Author:** Yusuf Çağrı İPİN, **E-mail:** cagriipin92@gmail.com

INTRODUCTION

The concept of sport is defined as the transformation of people's inherent desire to struggle and win into a regular and systematic competition (Kasap, 1997). Sports are conducted in one of two ways: amateur or professional, depending on the purpose of the sport in question. It is possible to talk about amateur sports when it is conducted for the purpose of having fun or providing physical development, and professional sports when it is conducted as a profession in which the individuals involved make financial gain. In both forms, the basic elements of sport consist of the concepts of athlete, spectator and referee. If there is another element among them that is at least as important as the athletes, it is the referees who make efforts together with the athletes in the field of play and affect the outcome of the game in a sense with their evaluations (Atılgan &Tükel, 2019).

Referee is defined as a person who manages sports competitions in accordance with the rules, determines the numbers or points won and is responsible for punishing those who do not comply with the rules of the game (Çelik, 2020; Durna, 1997). Refereeing involves a multifaceted and complex process that includes physical, mental and psychological dimensions. In some competitions, referees have to make crucial decisions in very short periods of time. Sport participants who desire to win expect referees to make correct and consistent decisions in such situations and throughout the competition (Eroğlu, 2018). Referees, who endeavour to manage an equal and fair game, have to struggle with multiple stimuli both on and off the field. In this context, the psychological and mental states of the referees, who should have some psychological, physiological and mental characteristics, may directly affect their evaluations and decisions (Ünsal, 2005).

Referees observe and evaluate the performance of athletes in individual sports competitions, while in team sports competitions, they are in the position of the person who makes decisions in order to apply the rules of the game correctly and impartially. At this point, it can be said that referees perform a measurement and evaluation process related to the performance of the athletes and the rules of the game in the competitions and competitions in which they take part. Since important decisions can be made about the athlete as a result of performance evaluation in sports environments, the measurement and evaluation is expected to be accurate and impartial, but in some cases it can be seen that referees make measurement and evaluation errors due to various factors (Çelik, 2020). There are many sources of error encountered in performance evaluations, including factors such as the measurement tool, the measurement environment, the measurement method, the person making the measurement and the situation of the rater at the time of measurement. Halo effect is defined as a kind of measurement error arising from the person making the measurement and the concept is widely studied in the field of performance and personality evaluation (Özgüven, 2007).

Halo effect is defined as a logical error that causes the evaluator to make positive or negative generalisations about all other qualities of an individual by being influenced by a prominent quality of the individual (Borman, 1975; Murphy et al., 1993). Halo effect, which is also defined as a type of cognitive prejudice, can cause us to perceive a person as a criminal or a character prone to committing a crime just because of his/her bad appearance (Nufer & Alesi, 2018).

The existence of the term, which is expressed in different ways in the literature such as "Halo Effect", "Ayla Effect" (Kağıtçıbaşı, 2010), "Generalisation Error" (Özgüven, 2007), "Logical Error" (Newcomb, 1931), "Imaginary Halo" (Cooper, 1981), was first put forward by the American psychologist Frederick L. Wells in 1907. It was officially defined in 1920 thanks to the experimental evidence provided by psychologist Edward Thorndike (Karakoç-Alatlı, 2012). Thorndike (1920) defines the concept of Halo effect as the effect of the first impression caused by a distinctive feature of a person, generalising to all the features of that person and affecting the decisions to be made about that person (Pekcan, 2019). Thorndike, in his research on how army commanders evaluate their soldiers, noticed that there is usually a high correlation between these ratings when different and unrelated characteristics of soldiers are evaluated. According to Thorndike's (1920) findings, it was observed that a soldier with an athletic and strong physical structure was also perceived as a soldier with intelligence, character and leadership qualities (Thorndike, 1920).

The word halo is defined in the English dictionary as "a ring of light around or above the head of a holy person in a religious drawing or painting" (Cambridge Dictionary, t.y.). It is stated that Thorndike (1920) named the concept in this way to depict how the Halo effect affects our perceptions in order to show that we assume that all other qualities of a person are also good based on a positive feature of the person, that is, we find that person worthy of the sacred ring of light named Halo (Suveren, 2022).

Halo effect is defined as a logic error that leads the evaluator to rate similar features in the same way (Newcomb, 1931). For example, the fact that a student who obeys and respects the rules is evaluated as a successful student by his/her teacher and his/her academic grade is rated high indicates the presence of Halo effect in the evaluation. This behaviour of the teacher shows that he/she evaluates by being under the influence of the student's general impression and generalising this impression to all other characteristics of the student (Karakoç-Alatlı, 2012).

People make many evaluations and judgements in daily life and make decisions as a result of them. Although they think that their decisions are objective and rational, the reality is that people's thoughts can often be unconsciously influenced by cognitive biases (Lance et al., 1994). Due to the large amount of information available in our environment but limited time to make decisions, it becomes impossible to fully process and analyse every piece of information that reaches our minds (Nufer, 2019). Therefore, people tend to use mental shortcuts that help to make quick and easy decisions. Although these shortcuts involving heuristics are generally thought to lead to accurate and valid results, they are also known to be highly susceptible to cognitive biases (Nufer & Alesi, 2018). The halo effect occurs because individuals' social perception consists of multiple factors and is a complex process. While forming impressions and making judgements about others, we do not rely solely on objective information; instead, we endeavour to create an image in our minds that fits the information we have previously acquired and interpret it in a way that our minds approve. Although they are surprisingly unrelated, the moments when we sometimes judge a person's character based on that person's physical attractiveness provide us with evidence that the Halo effect works (Nufer, 2018).

Cognitive biases known as perceptual illusions are quite common in daily life, and the Halo effect is one of the most common among them. Kozlowski, Kirsch & Chao (1986) state that the Halo effect can exist in almost all evaluations made by people, Pike (1999) states that the Halo effect can be seen continuously, and Cooper (1981) states that the Halo effect can occur at any time and in any situation and is an inevitable effect (Karakoç-Alatlı, 2012). The halo effect is a cognitive bias in which the first positive judgement about a person unconsciously leads to the perception of the individual as a whole. While forming the first impression of a person or object, observation of an initially attractive feature such as power or beauty may make that person or object attractive and may make it difficult to revise this impression even if new contrary information is obtained (Rogers, 2005). For example, evaluating an attractive individual as interesting, intelligent or funny at the same time indicates that this evaluation is made under the Halo effect, whether this evaluation is justified or not (Nufer, 2019).

Although different definitions of the concept have emerged as a result of the research on the halo effect, there are common views on the fact that it is an evaluator-induced error and is widely seen (Nisbett & Willson, 1977; Sigall & Ostrove, 1975). The Halo effect, sometimes known as the "physical attractiveness stereotype" (Palmer & Peterson, 2016) or the principle of "what is beautiful is also good" (Landy & Sigall, 1974), can be said to arise when we like a person, we usually assume that other characteristics of that person that we know less about are also positive (Yavuz, 2022).

The sport environment is an area where psychomotor skills are measured and evaluated, and it involves a complex process that requires referees to focus on more than one criterion when evaluating (Turgut, 1993). When important decisions are to be made as a result of this evaluation, it is of great importance that the scoring is done accurately and fairly. In addition, it can be observed that these evaluations sometimes produce biased or erroneous results (Arsan, 2012). From this point of view, are the referees affected by the distinctive characteristics of the athletes in the performance evaluations made in sports competitions and competitions? Does being under the influence of this prominent feature cause biased or erroneous evaluations? Are the biased or erroneous evaluations caused by the Halo effect? These questions constitute the basic questions that the researcher is curious about the subject and that led to the emergence of this research. By seeking answers to these questions with the scale developed within the scope of our research, it is aimed to find out whether there are clues pointing to the presence of Halo effect in referee evaluations in sports.

When the literature on sports refereeing was examined, it was seen that the researches mostly focused on decision making, decision making styles, communication skills, self-efficacy, moral attitudes, refereeing and its problems, duties, powers and responsibilities of referees, prejudice, referee bias, home advantage and favouritism. When the halo effect literature is analysed, it is seen that it is generally focused on personality and performance evaluation, measurement and evaluation in the field of education, marketing and sports sponsorship. However, there is no study in the literature in which the Halo effect is associated with referee evaluations in sport. In this direction, investigating the subject for the first time and aiming to develop a scale on the subject reveals the original value of our research. In addition, the results of the research are expected to contribute to the field and future studies to be conducted on this subject.

METHOD

Research Model

The aim of this study is to develop a scale that can measure the relationship between referees' evaluations of athlete performances and the Halo effect. Halo effect is a concept that has been little researched in sport environments. For this reason, problem statements or hypotheses were not developed for our exploratory study. Regarding the scale development study, the main sources on this subject and the scale development studies conducted on subjects similar to this study were utilised.

Research Group

The population of the study consists of referees in various sports branches and in different classifications, who are registered for the year 2023. While determining the sample number for this non-homogeneous population, a sample calculation (Özdamar, 2003) was made and 247 referees were reached in the study. In scale development studies, it is recommended to work 4 or 5 times the number of items (Hair et al., 1998). In this context, it was ensured that the study group of the research exceeded five times the number of items (46*5=230).

Ethical Approval

With the decision of Muğla Sıtkı Koçman University Social and Human Sciences Research Ethics Committee dated 29/11/2022 and numbered 220174/147, it was decided that there was no drawback in terms of scientific research ethics regarding the applicability of the research.

Data Collection

Scale Development Process

It is important that the construct to be measured has a theoretical basis and originality or generalisability. Theorising provides a great convenience in clarification. When there is no theory to guide, the researcher should prepare his/her own conceptual framework (DeVellis, 2003). In this direction, the concept of Halo effect was clearly defined through a comprehensive literature review, and its domain and boundaries were determined. Examining the models related to the concept of halo effect, it was seen that this effect can be encountered in almost every age, in every environment and in every situation, and its effect in sports environments was wanted to be investigated. While creating the item pool, all previous and accessible studies on the Halo effect were analysed. Among these studies, attention was paid especially to the studies that examined the concept of Halo effect by associating it with sports environments. The items produced by the researcher on the basis of the Halo effect models in the literature were constructed by taking into account the situations encountered in referee evaluations in sports and written as the referees' own statements. A total of 46 items were formed by taking into account each dimension believed to be the result of prejudiced attitudes in the events seen in referee decisions and referee evaluations in sports. In order to ensure the content validity of the items, a total of 7 experts in the fields of sport sciences, grammar and measurement and evaluation were consulted for their opinions on the items, and necessary arrangements and simplifications were made. After the expert opinions, a pilot study was conducted with the 46-item application form. The pilot application of the study was applied online with the 46-item pre-test form to a group of 43 referees in line with the recommendations of the field experts

(Şeker & Gençdoğan, 2020). As a result of the pilot application, it was seen that the internal validity of the scale was ensured, it was determined that there were no items to be removed and the actual application was started.

Data Analysis

The data obtained from the research were analysed using SPSS 22.0 and AMOS software. In these analyses, psychometric properties of the items such as item discrimination power and item total score correlations and psychometric properties of the scale such as construct validity and reliability measurements were tried to be determined.

FINDINGS

Seven expert opinions were consulted regarding content validity and content validity was calculated both at item level (CVR) and scale level (CVI) using Lawshe method. The content validity index was found to be 0.925 and the content validity ratios were found between 0.714-1 (Table 1).

Table 1. Content validity results of the scale items

	Expert 1	**Expert 2**	**Expert 3**	**Expert 4**	**Expert 5**	**Expert 6**	**Expert 7**	**CVR (KGO)**
Item 1	x	x	x	x	x	x	x	1
Item 2	x	x	x	x	x	x	x	1
Item 3	x	x	x	x	x	x	x	1
Item 4	x	x	x	x	x	x	x	1
Item 5	x	x	x	x	x	x	x	1
Item 6	x	x	x	x	x	x	x	1
Item 7	x	x	x	x	x	x	x	1
Item 8	x	x	x	x	x	x		0.714
Item 9	x	x	x	x	x	x	x	1
Item 10	x	x	x	x	x	x	x	1
Item 11	x	x	x	x		x	x	0.714
Item 12	x	x	x	x		x	x	0.714
Item 13	x	x	x	x		x	x	0.714
Item 14	x	x	x	x	x	x	x	1
Item 15	x	x	x	x		x	x	0.714
Item 16	x	x	x	x	x	x	x	1
Item 17	x	x	x	x	x	x	x	1
Item 18	x	x	x	x	x	x	x	1
Item 19	x	x	x	x	x	x	x	1
Item 20	x	x	x	x	x	x	x	1
Item 21	x	x	x	x		x	x	0.714
Item 22	x	x	x	x	x	x	x	1
Item 23	x	x	x	x	x	x	x	1
Item 24	x	x	x	x	x	x	x	1
Item 25	x	x	x	x	x	x	x	1
Item 26	x	x	x	x	x	x	x	1
Item 27	x	x	x	x	x	x	x	1
Item 28	x	x	x	x	x	x	x	1
Item 29	x	x	x	x	x	x	x	1
Item 30	x	x	x	x		x	x	0.714
Item 31	x	x	x	x	x	x	x	1
Item 32	x	x	x	x	x	x	x	1
Item 33	x	x	x	x		x	x	0.714

Table 1 (Continue). Content validity results of the scale items

	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	CVR (KGO)
Item 34	x	x	x	x	x	x	x	1
Item 35	x	x	x	x		x	x	0.714
Item 36	x	x	x	x	x	x	x	1
Item 37	x	x	x	x	x	x	x	1
Item 38	x	x	x	x	x	x	x	1
Item 39	x	x	x		x	x	x	0.714
Item 40		x	x	x	x	x	x	0.714
Item 41	x	x	x	x		x	x	0.714
Item 42	x	x	x	x	x	x	x	1
Item 43	x	x	x	x	x	x	x	1
Item 44	x	x	x	x	x	x	x	1
Item 45	x	x	x	x	x	x	x	1
Item 46	x	x	x	x	x	x	x	1
							CVI	**0.925**

As a result of the pilot study, reliability and item analysis were applied to the scale and the Alpha coefficient was found to be 0.970. As a result of the item analysis for the effect of the items on internal consistency, it was understood that the correlation coefficient with the total of the items was 0.3 and above, and it was determined that there were no items to be removed (Table 2).

Table 2. Pilot application reliability and item analysis

	Scale score when the item is deleted	Variance when item is deleted	Item total correlation	Cronbach alpha when the item is deleted
HALO1	77.750	930.238	.577	.970
HALO2	77.841	940.323	.560	.970
HALO3	75.318	918.362	.394	.971
HALO4	77.818	938.292	.612	.970
HALO5	77.636	920.283	.688	.969
HALO6	77.000	895.581	.677	.970
HALO7	77.568	914.298	.749	.969
HALO8	77.636	929.586	.562	.970
HALO9	76.773	908.645	.554	.970
HALO10	77.386	904.243	.710	.969
HALO11	75.659	917.765	.442	.971
HALO12	77.705	922.864	.705	.970
HALO13	77.705	926.911	.612	.970
HALO14	76.841	918.044	.434	.971
HALO15	77.068	924.344	.402	.971
HALO16	77.318	897.710	.798	.969
HALO17	77.295	904.306	.730	.969
HALO18	77.500	908.256	.836	.969
HALO19	77.273	913.552	.695	.969
HALO20	77.682	923.943	.786	.969
HALO21	77.455	909.416	.698	.969
HALO22	77.500	910.116	.757	.969
HALO23	77.136	897.981	.824	.969
HALO24	77.136	903.423	.727	.969
HALO25	76.614	906.196	.627	.970

Table 2 (Continue). Pilot application reliability and item analysis

	Scale score when the item is deleted	Variance when item is deleted	Item total correlation	Cronbach alpha when the item is deleted
HALO26	76.773	906.459	.650	.970
HALO27	77.227	918.040	.539	.970
HALO28	76.955	905.672	.640	.970
HALO29	77.636	925.865	.720	.970
HALO30	77.659	928.323	.627	.970
HALO31	77.273	903.784	.772	.969
HALO32	77.818	937.780	.472	.970
HALO33	77.432	905.879	.774	.969
HALO34	77.318	930.501	.356	.971
HALO35	77.386	910.522	.775	.969
HALO36	77.568	911.088	.839	.969
HALO37	77.477	903.790	.848	.969
HALO38	77.023	904.720	.682	.969
HALO39	77.250	928.099	.419	.970
HALO40	77.432	920.716	.643	.970
HALO41	77.136	897.330	.789	.969
HALO42	77.273	895.645	.850	.969
HALO43	77.023	903.186	.680	.969
HALO44	77.545	911.323	.781	.969
HALO45	76.864	905.423	.616	.970
HALO46	77.295	909.236	.668	.969

Reliability analysis was performed for the scale and Alpha coefficient was found as 0.972. When the item analysis was analysed, it was found that the item total correlation value of item number 3 was below 0.3 (.278) and it was removed from the scale (Table 3).

Table 3. Main application reliability and item analysis

	Scale score when the item is deleted	Variance when item is deleted	Item total correlation	Cronbach alpha when the item is deleted
HALO1	83.06	1016.740	.410	.972
HALO2	83.12	1017.925	.458	.972
HALO3	84.39	1036.483	.278	.972
HALO4	83.13	1023.186	.382	.972
HALO5	82.92	999.278	.691	.972
HALO6	82.36	972.019	.788	.971
HALO7	82.76	989.071	.773	.971
HALO8	82.55	1006.021	.392	.973
HALO9	81.88	981.042	.633	.972
HALO10	82.42	972.294	.805	.971
HALO11	81.31	986.663	.501	.973
HALO12	82.92	1001.896	.645	.972
HALO13	83.15	1019.171	.464	.972
HALO14	82.50	992.178	.537	.972
HALO15	82.28	993.469	.489	.973
HALO16	82.53	973.933	.838	.971
HALO17	82.45	969.996	.852	.971
HALO18	82.94	1002.972	.621	.972
HALO19	82.74	1000.093	.620	.972
HALO20	82.97	1002.166	.734	.972
HALO21	82.70	985.331	.814	.971
HALO22	82.77	984.284	.830	.971
HALO23	82.23	970.249	.781	.971

Table 3 (Continue). Main application reliability and item analysis

	Scale score when the item is deleted	Variance when item is deleted	Item total correlation	Cronbach alpha when the item is deleted
HALO25	81.50	982.601	.644	.972
HALO26	81.40	986.874	.627	.972
HALO27	82.75	1002.114	.609	.972
HALO28	82.23	994.560	.587	.972
HALO29	82.85	1017.290	.400	.972
HALO30	82.91	1004.971	.684	.972
HALO31	82.16	971.787	.791	.971
HALO32	83.23	1025.802	.441	.972
HALO33	82.22	981.106	.720	.971
HALO34	82.75	1006.538	.481	.972
HALO35	82.50	973.926	.834	.971
HALO36	82.61	980.890	.825	.971
HALO37	82.60	989.183	.725	.972
HALO38	81.77	981.888	.679	.972
HALO39	82.87	1007.785	.548	.972
HALO40	83.00	1011.102	.590	.972
HALO41	82.18	960.385	.854	.971
HALO42	82.36	972.541	.827	.971
HALO43	82.39	973.947	.833	.971
HALO44	82.85	992.608	.811	.971
HALO45	82.08	981.681	.636	.972
HALO46	82.45	977.834	.760	.971

Exploratory factor analysis method was applied to reveal the construct validity of the scale. As a result of Barlett sphericity test ($p=0.000<0.05$) and Kaiser-Meyer-Olkin test ($KMO=0.908>0.60$), it was determined that the sample size was sufficient for factor analysis. As a result of the factor analysis, the variables were grouped under 3 factors with a total explained variance of 72.321%.

Items numbered 1,2,4,8,9,11,12,13,15,19,20,29,32,34,38,39,45,46 were removed from the scale since their co-loading and factor loading were below 0.3. In order to calculate the reliability of the remaining 27 items in the scale, the internal consistency coefficient "Cronbach Alpha" was calculated. The overall reliability of the scale was found as Alpha=0.972. The 12 items in the first factor were named as "Distinct Size Effect", 9 items in the second dimension were named as "Insufficient Discrimination Effect" and 6 items in the third dimension were named as "General Impression Effect" (Table 4).

Table 4. Explanatory factor analysis results

Dimension	Factor Load
Distinct Dimension Effect (Eigenvalue=16.035; Variance Explained=30.067; Alpha=0.964)	
10. The fact that the athlete is a citizen of a country that is friendly with my country has a positive effect on my assessment of him/her.	0.829
21. I look favourably on athletes from clubs and countries with strong managers.	0.805
16. The high prestige of the athlete's country in the international arena has a positive influence on my evaluation of the athlete.	0.801
14. I want athletes with the same religion as me to win the competition.	0.790
17. The fact that the athlete's country is at the top of the world rankings in that sport has a positive effect on my evaluation of the athlete.	0.789
22. I have a favourable view of athletes of the same nationality as important people in international federations.	0.759
43. I look favourably on athletes I see regularly in sports competitions.	0.723
41. The fact that the athlete has previously won European, World or Olympic honours in the relevant discipline influences my assessment of him/her positively.	0.687
6. Being a citizen of my own country has a positive effect on my assessment of the athlete.	0.668
42. When an athlete whom I like in the media participates in a competition, I make positive evaluations about that athlete.	0.663
23. I consider athletes from strong countries and clubs as favourites in competitions.	0.642
37. The fact that I have a bad memory with the athlete negatively affects my evaluation of him/her.	0.480
Insufficient Discrimination Effect (Eigenvalue=1.906; Variance Explained=21.940; Alpha=0.930)	
5. The fact that an athlete is from a nation I don't like negatively affects my evaluation of him/her.	0.799
18. The fact that the athlete is a citizen of an undeveloped country has a negative impact on my assessment of him/her.	0.798
35. The fact that the athlete is a famous person in that sport has a positive effect on my evaluations about him/her.	0.682
7. The fact that the athlete is a citizen of a country whose diplomatic interests do not coincide with those of my country may have a negative impact on my assessment of him/her.	0.673
30. Whether an athlete is beautiful or handsome has a positive effect on my evaluation of him/her.	0.637
44. An athlete's host status in a competition has a positive effect on my assessment of him/her.	0.625
27. The athlete's impressive physique has a positive effect on my evaluation of him/her.	0.612
40. I consider an athlete who is not recognised in a sport to be an unsuccessful athlete.	0.575
36. Meeting the athlete has a positive effect on my evaluation of him/her.	0.569
General Impression Effect (Eigenvalue=1.586; Variance Explained=20.314; Alpha=0.906)	
25. An athlete's rude behaviour has a negative impact on my assessment of him/her.	0.855
26. If an athlete has a disrespectful way of speaking, it has a negative impact on my judgement of him/her.	0.817
24. If the athlete uses gestures and facial expressions that I do not like, it affects my evaluation of him/her negatively.	0.686
31. If an athlete is arrogant, it has a negative effect on my judgement of him.	0.670
33. Exaggerated behaviour of the athlete has a negative effect on my evaluation of him/her.	0.666
28. I am attracted to athletes with a sympathetic demeanour.	0.555
Total Variance=72.321%; Overall Reliability (Alpha)=0.972	

In the study, the most frequently used goodness of fit indices in the literature were used. The diagram of confirmatory factor analysis is given below (Figure 1).

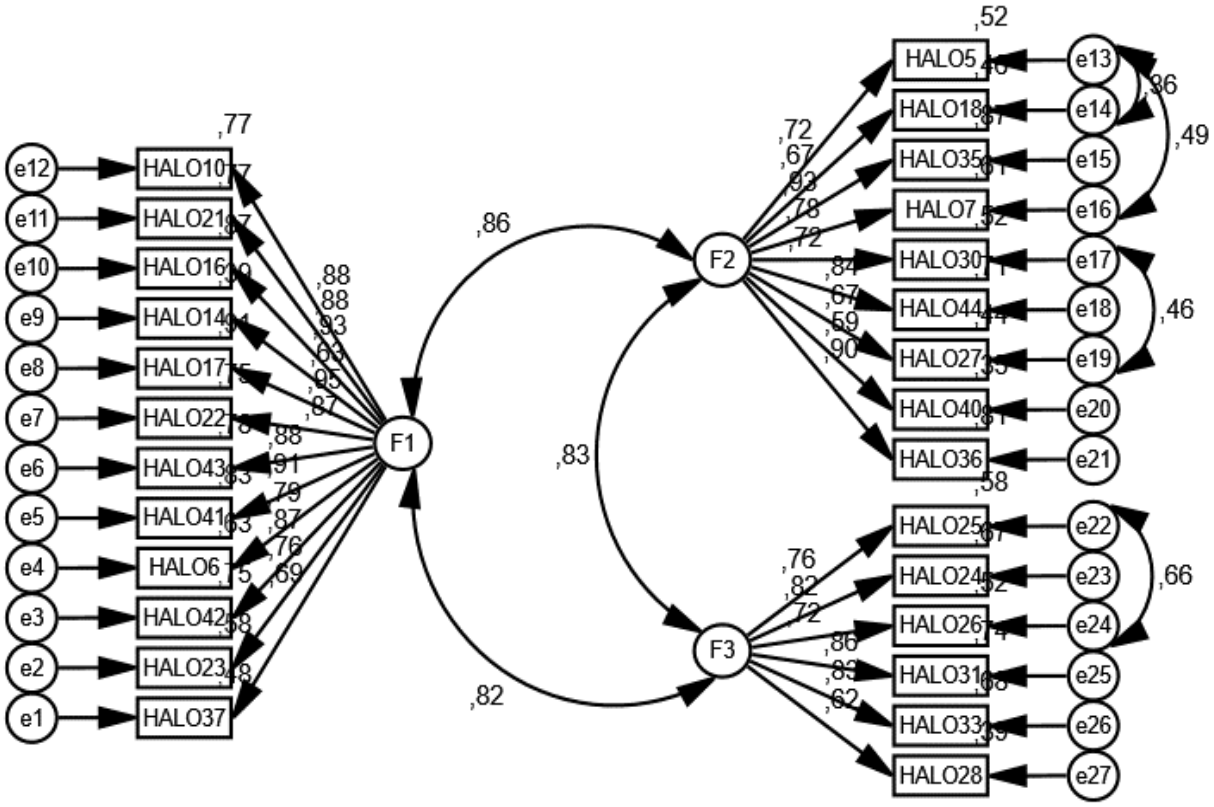

Figure 1. Diagram for confirmatory factor analysis

In the results of the analyses, it was found that the fit statistics calculated by confirmatory factor analysis were compatible with the previously determined factor structure of the scale (Table 5).

Table 5. Confirmatory factor analysis index values

Index	Normal value	Acceptable value	Value found
$\chi 2$/sd	<2	<5	3.69
GFI	>0.95	>0.90	0.90
AGFI	>0.95	>0.90	0.90
CFI	>0.95	>0.90	0.91
RMSEA	<0.05	<0.08	0.06
RMR	<0.05	<0.08	0.07

When factor loadings related to confirmatory factor analysis were examined, it was found that the factor loadings were high, standard error values were low and t values were significant (Table 6).

Table 6. Confirmatory factor analysis factor loadings

Items	Factors	β	Std. β	S.Error	t	p
HALO37	F1	1.000	.691			
HALO23	F1	1.437	.760	.125	11.476	p<0.001
HALO42	F1	1.500	.868	.115	13.009	p<0.001
HALO6	F1	1.447	.793	.121	11.950	p<0.001
HALO41	F1	1.822	.912	.134	13.612	p<0.001
HALO43	F1	1.478	.881	.112	13.191	p<0.001
HALO22	F1	1.218	.868	.094	13.008	p<0.001
HALO17	F1	1.663	.952	.117	14.164	p<0.001
HALO14	F1	1.128	.626	.118	9.537	p<0.001
HALO16	F1	1.556	.932	.112	13.899	p<0.001
HALO21	F1	1.232	.880	.094	13.176	p<0.001
HALO10	F1	1.560	.877	.119	13.130	p<0.001
HALO5	F2	1.000	.720			
HALO18	F2	.952	.675	.073	13.071	p<0.001
HALO35	F2	1.859	.934	.127	14.671	p<0.001
HALO7	F2	1.260	.779	.075	16.886	p<0.001
HALO30	F2	.872	.721	.078	11.225	p<0.001
HALO44	F2	1.200	.840	.091	13.148	p<0.001
HALO27	F2	.982	.666	.095	10.343	p<0.001
HALO40	F2	.668	.594	.073	9.204	p<0.001
HALO36	F2	1.605	.900	.114	14.119	p<0.001
HALO25	F3	1.000	.760			
HALO24	F3	.974	.817	.073	13.352	p<0.001
HALO26	F3	.897	.721	.045	19.758	p<0.001
HALO31	F3	1.121	.862	.079	14.203	p<0.001
HALO33	F3	1.006	.826	.074	13.538	p<0.001
HALO28	F3	.699	.625	.071	9.864	p<0.001

When the discrimination analysis of the scale was analysed, it was found that the scale showed a significant difference (p<0.05) between the lower 27% and upper 27% groups (Table 7).

Table 7. Results of the discrimination analysis

Groups	Lower %27 (n=67)		Upper %27 (n=67)		t	sd	p
	Avg.	S	Avg.	S			
Halo Effect General	1.100	0.078	3.105	0.537	-30.236	132	**0.000**
Distinct Dimensional Effect	1.086	0.113	3.340	0.742	-24.591	132	**0.000**
Insufficient Discrimination Effect	1.020	0.072	2.438	0.753	-15.352	132	**0.000**
General Impression Effect	1.249	0.316	3.637	0.533	-31.540	132	**0.000**

When the convergent and divergent validity analyses of the scale were examined, it was found that CR (composite reliability) values were greater than AVE (average variance extracted) values and AVE values were greater than 0.5. The square root of the AVE value of each factor was found to be greater than the correlation values of that factor with other factors (Table 8).

Table 8. Convergent validity and divergent validity analysis results

	CR	AVE
Distinct Dimensional Effect	0.745	0.536
Insufficient Discrimination Effec	0.799	0.519
General Impression Effect	0.769	0.612

When the analysis performed to reveal the time-dependent measurement invariance of the scale was examined, it was found that the ICC values related to the agreement between the test-retest measurements were high, and there was no difference between the test-retest correlation values (p<0.05) (Table 9).

Table 9. Test-retest results

Measurements	Test		Re-Test		N	t	p[a]	ICC[b]	p
	Avg.	Ss	Avg.	Ss					
Halo Effect General	1.823	0.899	1.876	0.884	40	-1.646	0.108	0.945	0.000
Distinct Dimensional Effect	1.892	1.137	1.944	1.121	40	-1.602	0.117	0.936	0.000
Insufficient Discrimination Effec	1.458	0.696	1.508	0.698	40	-1.433	0.160	0.947	0.000
General Impression Effect	2.233	1.071	2.175	0.999	40	0.894	0.377	0.918	0.000

DISCUSSION AND CONCLUSION

In this study, in order to measure the relationship between referees' evaluations of athletes' performances and the Halo effect, a 27-item scale consisting of 3 sub-dimensions as " Distinct Dimension Effect", "Insufficient Discrimination Effect" and "General Impression Effect" was developed. "Distinct Dimension Effect" dimension measures the extent to which the referees make the same evaluations about different characteristics of the athletes under the influence of a prominent dominant characteristic and there are 12 items in this sub-dimension. "Insufficient Discrimination Effect" dimension measures the referees' inability to discriminate between the different and unrelated characteristics of the athletes while making evaluations, and there are 9 items in this sub-dimension. The "General Impression Effect" dimension measures the level of the referees' being under the influence of the general impression of the athletes while making evaluations and there are 6 items in this sub-dimension.

Content validity is used to determine the extent to which a scale and its items serve the purpose of measurement (Erkan & Gömleksiz, 2014). The content validity index was calculated at both item level (CVR) and scale level (CVI) by consulting 7 experts using the method developed by Lawshe (1975). According to Lawshe, when the CVR of 7 expert opinions is 0.75 or greater, the consensus among experts is statistically valid (Romero-Jeldres et al., 2023). The content validity index for the scale was found to be 0.925 and the content validity ratios were found between 0.714-1. This result shows that the content validity of the items in the scale is sufficient.

To test the reliability of the scale, the item-total score correlation analysis method was applied. Item analysis is a reliability technique that determines the relationship between each item in the scale and the measured structure and the total score of the scale (Gökdemir & Yılmaz, 2023). The responses to the items are expected to have a positive correlation between the items and with the total scale. This shows that the participants understand the propositions correctly and respond objectively. The fact that the correlation coefficient of an item in the scale with the sum of items is 0.3 and above indicates that the discrimination of the scale is high (Büyüköztürk, 2011; Tavşancıl, 2018). When the reliability and item analyses conducted after the pilot application were examined, it was found that there were no items with item-total correlation values below 0.3 and no items that negatively affected the internal consistency. When the reliability and item analyses conducted after the actual implementation were examined, it was found that the item total correlation value of item number 3 was below 0.3 (.278) and it was removed from the scale. Reliability and item analysis were repeated and the Alpha coefficient was found to be 0.972.

As a result of KMO (KMO=0.908>0.60) and Bartlett's test (p=0.000<0.05), it was found that the data set was suitable for factor analysis and exploratory factor analysis (EFA) method was applied to reveal the construct validity of the scale. In factor analysis, the correlation coefficient determines the relationship between two variables. High values for the correlation coefficient allow a factor to form, while low values indicate a weak relationship between variables and prevent factor formation. In this context, it is recommended in the literature that the correlation coefficient should be at least .32 and that variables with lower correlations should be excluded from factor analysis (Sürücü et al., 2024). In this context, a total of 19 items were removed from the scale because their co-loadings and factor loadings were below 0.3. In order to calculate the reliability of the remaining 27 items in the scale, the internal consistency coefficient "Cronbach Alpha" was calculated. It is stated in the literature that the internal consistency coefficient should be at least .70 and above (Çokluk et al., 2021). Accordingly; the high Alpha coefficients for the sub-dimensions of the scale (Distinct Dimension Effect = .964, Insufficient Discrimination Effect = .930 and General Impression Effect = .906) indicate that the items in the sub-dimensions are consistent with each other. The overall reliability of the scale was found as Alpha=0.972. Based on the findings obtained from the EFA analysis, it was understood that the scale was a valid and reliable measurement tool.

Confirmatory Factor Analysis (CFA) is a type of structural equation model (SEM) that can measure the relationship between observed variables and latent variables. For the confirmatory factor analysis (CFA) of the study, the most frequently used goodness of fit indices in the literature were used (Tabachnick & Fidell, 2001). A comparison is made between the basic model and the proposed model through these indices. For these indices, a value of 1 indicates that the proposed model fits the data perfectly according to the basic model, while a value of 0 means that the proposed model does not have any explanatory value (Goretzko et al., 2024). The fit statistics calculated by CFA analyses were found to be compatible with the previously determined factor structure of the scale at an acceptable level and the CFA results confirmed the validity of the scale.

In addition to item analysis, discrimination test was also applied. The discrimination test is the determination of the significant difference between the groups by dividing the total score of the scale into groups as Lower 27% and Upper 27%. A difference between two groups is an indicator of discrimination. No difference between the two groups indicates that the lowest and highest score range is small (Tezbaşaran, 2008). When this method is applied, individuals are divided into lower 27% and upper 27% groups from largest to smallest according to the total scores they receive from the scale and comparison is made with the t test. If statistically significant differences occur, it can be said that the items in the scale can distinguish individuals in the upper and lower groups in terms of the measured feature (Gökdemir & Yılmaz, 2023). As a result of the analysis, it was found that the scale showed a significant difference between the Lower 27% and Upper 27% groups (p<0.05). According to these results, it can be said that the scale makes sensitive measurements to distinguish.

In order to test the construct validity of the variables in the measurement model, Composite Reliability (CR) and Average Variance Explained (AVE) values were analysed. In order to ensure convergent validity, it is expected that the CR values of the scale should be greater than

the AVE values and the AVE value should be greater than 0.5 (Fornell & Larckers, 1981). A CR value of 0.7 and above is considered a strong value in the literature because it shows that a significant portion of the variance is explained by the structure (Lim, 2024). When the analysis results were examined, it was found that the AVE values were greater than 0.5 and the CR values were greater than 0.7. Accordingly, it is seen that convergent validity and discriminant validity of the scale are provided.

In addition, test-retest analyses were performed on 40 participants in order to reveal time-dependent measurement invariance. In reliability studies, ICC (Intraclass Correlation Coefficient) values close to 1 indicate that the test is reliable (Mondal et al., 2024). It was found that ICC values related to the agreement between test-retest measurements were high (0.918-0.947) and there was no difference between test-retest correlation values ($p<0.05$). According to this finding, it can be said that the scale makes reliable measurements based on short time (Shrout & Fleiss, 1979).

As a result, in line with the findings obtained within the scope of the research, it was concluded that the "Perceived Halo Effect Scale in Referee Evaluations in Sports" is a valid and reliable measurement tool. With this scale study, it was aimed to create awareness about the concept of prejudice in referees and to raise awareness and inform the referees about personal prejudices. In this way, it will be ensured that the referees are aware of the decisions they make or the evaluations they make in the sports competitions they manage and that they are aware of whether their personal attitudes and prejudices affect their evaluations. Thanks to this awareness, it is thought that the number of referees who can act morally and ethically and make objective evaluations as required by the refereeing identity will increase.

SUGGESTIONS

The application of future studies on the subject on a larger population and samples may provide more meaningful findings. Referees make decisions within the rules of the game in team sports. In individual sports, they make subjective evaluations about athletes. For this reason, focusing on referees working in sports branches where individual evaluations are made may provide clearer clues on the subject. Since biased attitudes seen in referee evaluations are more common in international sports competitions, future studies can be applied on referees working in the international classification. In order to reveal cultural differences on a global scale, future studies can be applied to different cultures.

Conflict of Interest: The authors of the article have no personal or financial conflict of interest within the scope of the study.

Declaration of Contribution Rate of Researchers: Research design- YÇİ; HÜ, Data collection- YÇİ, Statistical analysis- YÇİ; HÜ, Manuscript preparation- YÇİ; HÜ.

Ethical Aprovval
Board Name: Muğla Sıtkı Koçman University Social and Humanities Research Ethics Committee
Date: 29/11/2022
Issue Number: 220174/147

REFERENCES

Arsan, N. (2012). *Buz pateninde hakem değerlendirmelerinin genellenebilirlik kuramı ve rasch modeli ile incelenmesi*. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü Doktora Tezi, Ankara.

Atılgan, D., & Tükel, Y. (2019). Hakemlerin karar verme stillerinin incelenmesi. *Sport Sciences, 14*(2), 22-32. http://dx.doi.org/10.12739/NWSA.2019.14.2.2B0119

Borman, W. C. (1975). Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. *Journal of Applied Psychology, 60*(5), 556- 560. https://doi.org/10.1037/0021-9010.60.5.556

Büyüköztürk, Ş. (2011). *Sosyal bilimler için veri analizi el kitabı*. Pegem Yayıncılık.

Cambridge Dictionary. (t.y.). Halo. in English dictionary. Retrieved May 5, 2023, from https://dictionary.cambridge.org/dictionary/english/halo

Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin, 90*(1), 218-244. https://doi.org/10.1037/0033-2909.90.2.218

Çelik, D. Ö. (2020). Farklı spor dallarında görev yapan hakemlerin karar verme stillerinin incelenmesi. *Gaziantep Üniversitesi Spor Bilimleri Dergisi, 5*(4), 385-398. https://doi.org/10.31680/gaunjss.776473

Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2021). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve Lisrel uygulamaları* (6. basım). Pegem Akademi.

DeVellis, R. F. (2003). *Scale development theory and applications.* SAGE Publication.

Durna, E. (1997). *Türkiye'de futbol ve hakem.* Yıldızlar Matbaacılık.

Erkan, S., & Gömleksiz, M. (2014). *Eğitimde ölçme ve değerlendirme.* Anı Yayıncılık.

Eroğlu, S. Y. (2018). *Hakemlerin ahlaki olgunluk ve empati düzeylerinin karar verme stillerine etkisinin incelenmesi*. Kütahya Dumlupınar Üniversitesi Sosyal Bilimler Enstitüsü Doktora Tezi, Kütahya.

Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement errors. *Journal of Marketing Research, 18*(2), 39–50. https://doi.org/10.2307/3151312

Goretzko, D., Siemund, K., & Sterner, P. (2024). Evaluating model fit of measurement models in confirmatory factor analysis. *Educational and Psychological Measurement, 84*(1), 123-144. https://doi.org/10.1177/00131644231163813

Gökdemir, F., & Yılmaz, T. (2023). Likert tipi ölçekleri kullanma, modifiye etme, uyarlama ve geliştirme süreçleri. *Journal of Nursology, 26*(2), 148-160. https://doi.org/10.5152/JANHS.2023.22260

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis.* Macmillan Publishing Company.

Kağıtçıbaşı, Ç. (2010). *Günümüzde insan ve insanlar.* Evrim Yayınları.

Karakoç-Alatlı, B. (2012). *Başarının değerlendirilmesinde algılanan halo etkisi ölçeği'nin geliştirilmesi.* Ankara Üniversitesi Eğitim Bilimleri Enstitüsü Yüksek Lisans Tezi, Ankara.

Kasap, H. (1997). *Beden eğitimi.* Morpa Spor Ansiklopedisi, 1, 156-160.

Kozlowski, S. W. J., Kirsch, M. P., & Chao, G. T. (1986). Job knowledge, ratee familitarity and halo error: An exploration. *Journal of Applied Psychology, 71*(1), 45-49. https://doi.org/10.1037/0021-9010.71.1.45

Lance, C. E., LaPointe, J. A., & Stewart, A. M. (1994). A test of the context dependency of three causal models of halo rater error. *Journal of Applied Psychology, 79*(3), 332-340. https://doi.org/10.1037/0021-9010.79.3.332

Landy, D., & Sigall, H. (1974). Beauty is talent: Task evaluation as a function of the performer's physical attractiveness. *Journal of Personality and Social Psychology*, *29*(3), 299-304. https://doi.org/10.1037/h0036018

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Lim, W. M. (2024). A typology of validity: Content, face, convergent, discriminant, nomological and predictive validity. *Journal of Trade Science, 12*(3), 155-179. https://doi.org/10.1108/JTS-03-2024-0016

Mondal, D., Vanbelle, S., Cassese, A., & Candel, M. J. (2024). Review of sample size determination methods for the intraclass correlation coefficient in the one-way analysis of variance model. *Statistical Methods in Medical Research, 33*(3), 532-553. https://doi.org/10.1177/09622802231224657

Murphy, K. R., Jako, R. A., & Anhalt, R. L. (1993). Nature and consequences of halo error: A critical analysis. *Journal of Applied Psychology, 78*(2), 218–225. https://doi.org/10.1037/0021-9010.78.2.218

Newcomb, T. M. (1931). An experiment designed to test the validity of rating technique. *Journal of Educational Psychology, 22*, 279-289. https://doi.org/10.1037/h0070737

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology, 35*(4), 250-256. https://doi.org/10.1037/0022-3514.35.4.250

Nufer, G., & Alesi, N. (2018). The halo effect in sports. *International Journal of Business and Social Science, 9*(2), 31-44. https://doi.org/10.30845/ijbss

Nufer, G. (2018). An exploration of the halo effect in professional soccer. *European Journal of Physical Education and Sport Science, 4*(9), 17-29. https://zenodo.org/records/1319077

Nufer, G. (2019). Say hello to halo: The halo effect in sports. *Innovative Marketing, 15*(3), 116-129. https://doi.org/10.21511/im.15(3).2019.09

Özdamar, K. (2003). *Modern bilimsel araştırma yöntemleri.* Kaan Kitabevi.

Özgüven, İ. E. (2007). *Psikolojik testler.* PDREM Yayınları.

Palmer, C. L., & Peterson, R. D. (2016). Halo effects and the attractiveness premium in perceptions of political expertise. *American Politics Research*, *44*(2), 353– 382. https://doi.org/10.1177/1532673X15600517

Pekcan, N. (2019). *Algılanan halo etkisinin hiyerarşik doğrusal modelleme ile incelenmesi*. Ankara Üniversitesi Eğitim Bilimleri Enstitüsü Yüksek Lisans Tezi, Ankara.

Pike, G. R. (1999). The constant error of the halo in educational outcomes research. *Research in Higher Education*, *40*(1), 61-86. https://doi.org/10.1023/A:1018774311468

Rogers, E. P. (2005). *Scale construction and halo effect in secondary student ratings of teacher performance*. Brigham Young University Department of Instructional Psychology and Technology Doctoral Thesis, Canada.

Romero-Jeldres, M., Díaz-Costa, E., & Faouzi-Nadim, T. (2023). A review of Lawshe's method for calculating content validity in the social sciences. *Front., Educ., 8*,1271335. https://doi.org/10.3389/feduc.2023.1271335

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Sigall, H., & Ostrove, N. (1975). Beautiful but dangerous: Effects of offender attractiveness and nature of the crime on juridic judgment. *Journal of Personality and Social Psychology*, *31*(3), 410-414. https://doi.org/10.1037/h0076472

Suveren, Y. (2022). Unconscious bias: Definition and significance. *Current Approaches in Psychiatry*, *14*(3), 414-426. https://doi.org/10.18863/pgy.1026607

Sürücü, L., Yıkılmaz, İ., & Maşlakçı, A. (2024). Exploratory factor analysis (EFA) in quantitative researches and practical considerations. *Gümüşhane Üniversitesi Sağlık Bilimleri Dergisi, 13*(2), 947-965. https://doi.org/10.37989/gumussagbil.1183271

Şeker, H., & Gençdoğan, B. (2020). *Psikolojide ve eğitimde ölçme aracı geliştirme* (3. basım). Nobel Akademik Yayıncılık.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th Edition). Allyn and Bacon.

Tavşancıl, E. (2018). *Tutumların ölçülmesi ve SPSS ile veri analizi* (6. basım). Nobel Akademik Yayıncılık.

Tezbaşaran, A. A. (2008). *Likert tipi ölçek hazırlama kılavuzu*. Türk Psikologlar Derneği.

Thorndike, E. L. (1920). A constant error on psychological rating. *Journal of Applied Psychology*, *4*, 25-29. https://doi.org/10.1037/h0071663

Turgut, F. M. (1993). *Eğitimde ölçme ve değerlendirme metotları* (9. basım). Saydam Matbaacılık.

Ünsal, A. (2005). *Tribün cemaatinin öfkesi, ticarileşen Türkiye futbolunda şiddet*. İletişim Yayınları.

Yavuz, E. (2022). *Görünümü taklit eden markaların, orijinal marka imajına etkisi: Halo etkisi üzerine bir araştırma.* Bahçeşehir Üniversitesi, Lisansüstü Eğitim Enstitüsü, Yüksek Lisans Tezi, İstanbul.