



Program Değerlendirme Tezlerinin Çok Yüzeyli Rasch Ölçme Modeli ile Meta Değerlendirmesi

Hatice Cura Yeleğen*^a, Onur Rahman Horuz^b, Büşra Kanıkırmızı^c, Büşra GÜVENÇ^d,

Makale Bilgisi

DOI:

Makale Geçmişi:

Geliş :10.06.2022

Düzeltilme :16.06.2022

Kabul :27.06.2022

Keywords:

Program değerlendirme,

Rasch ölçme modeli,

Meta değerlendirme.

Makale Türü:

Araştırma Makalesi

Öz

Araştırmanın amacı, "Program değerlendirme tezlerinin Rasch ölçme modeli ile meta değerlendirmesi" nin gerçekleştirilmesidir. Bu amaç doğrultusunda; program değerlendirme tezlerinin kalibrasyon haritası, program değerlendirme tezlerinin analizi, jürilerin puanlama davranışlarına yönelik analizler, program değerlendirme ölçütlerinde bulunan maddelerin madde güçlüklerinin analizi ve jürilerin tarafsızlık analizi yapılmıştır. Bu çalışmada tarama modeli kullanılmıştır. Çalışma grubu, 2020-2021 bahar akademik yılında Yükseköğretim Kurulu tez taramasında program değerlendirme anahtar kelimesiyle bulunan 16 tezi değerlendiren Eğitim Programları ve Öğretim bilim dalında yüksek lisans ve doktora yapan toplam 5 öğrencidir. Araştırmada 5 jüri, 16 tez ve 12 kriter kullanılmıştır. Jüriler program değerlendirme alanındaki 16 tezin puanlamasını 12 kriter ile, "Hiç", "Az", "Kısmen", "Çok" ve "Çok fazla" derecelendirmelerini kullanarak gerçekleştirmişlerdir. Araştırma sonuçlarına göre, en nitelikli tezin tez12, en yetersiz nitelikli tezin tez13, en hoşgörülü jürinin J4, en sert jürinin de J3 olduğu görülmüştür. Ayrıca çalışmanın, değerlendirmede kullanılan program değerlendirme modeli ile uyumlu olmasının en zor sağlanan kriter olduğu sonucu elde edilmiştir. Bu sebeple, program değerlendirme ile ilgili yapılacak olan çalışmalarda program değerlendirme alanında eğitim almış bir kişinin bulunması veya program değerlendirme alanında eğitim almış bir kişiden görüş alınması önerilmektedir.

*İlgili Yazar: curahatice@gmail.com

^a Bartın Üniversitesi Eğitim Programları ve Öğretim Bilim Dalı, 0000-0002-0726-7491, Bartın, Türkiye

^b Bartın Üniversitesi Eğitim Programları ve Öğretim Bilim Dalı, 0000-0001-8115-0104, Bartın, Türkiye

^c Bartın Üniversitesi Eğitim Programları ve Öğretim Bilim Dalı, 0000-0002-1977-3850, Bartın, Türkiye

^d Bartın Üniversitesi Eğitim Programları ve Öğretim Bilim Dalı, 0000-0002-1981-4480, Bartın, Türkiye

Meta Evaluation of the Program Evaluation Theses with Multifaceted Rasch Measurement Model

Article Information

DOI:

Article History:

Received :10.06.2022

Revised :16.06.2022

Accepted : 27.06.2022

Keywords:

Program Evaluation,
Rasch measurement
model,
Meta evaluation.

Article Type:

Research Article

Abstract

The aim of the research is to carry out the meta-evaluation of the program evaluation theses with the Rasch measurement model. In accordance with this purpose; The calibration map of the program evaluation theses, the analysis of the program evaluation theses, analysis of raters' rigidity or generosity., the analysis of the item difficulties of program evaluation criteria and and raters' bias analysis were performed. Survey model was used in the research. The study group consists of a total of 5 graduate and doctorate students in the field of Education Programs and Instruction, who evaluated 16 theses found with the keyword program evaluation in the Higher Education Council thesis screening in the 2020-2021 spring academic year. In the research, 5 raters, 16 theses and 12 criteria were used. According to the results of the research, it was seen that the most qualified thesis was thesis12, the lowest qualified thesis13, the most generous rater was P4, and the most rigorous was P3. In addition, it was concluded that the study was the most difficult criterion to be compatible with the program evaluation model used in the evaluation.

Giriş

Eğitim programının geliştirilmesine imkân sağlamak için yapılan program değerlendirme (Erden, 1998), program geliştirme sürecinin son ve tamamlayıcı basamağı (Ertürk, 2017) olarak programın etkililiği hakkında bir karara varmaya olanak tanıyan bir süreçtir (Demirel, 2020). Program değerlendirme, bahsedilen bu sürecin sonucunda program geliştirme uzmanlarına; programa devam edilmesi, programın tekrar ele alınarak incelenmesi ya da yeni bir aşamaya geçilmesi konusunda bilgi verir (Demirel, 2020).

Program değerlendirme çalışmaları belli bir sistematığe uyularak gerçekleştirildiğinde değerlendirmenin yapılaşdırılmasına ve daha kapsamlı hale getirilmesine olanak tanımaktadır (Kaya, 1997). Bir eğitim programını değerlendirirken sistematikleştirilmiş bilimsel bir araştırma süreci yürütülür ve bu süreç on iki temel aşamada ele alınabilmektedir (Uşun, 2016):

1. Program değerlendirmeyle alakalı problemin farkına varılması ve değerlendirme araştırmasıyla ilgili konunun belirlenmesi.
2. Program değerlendirmeyle alakalı amaç ve araştırma probleminin belirlenmesi.
3. Değerlendirme araştırmasıyla ilgili süre, imkân ve sınırlılıkların belirlenmesi.
4. İç ve dış değerlendirmecilerin belirlenmesi.
5. Konuyla alakalı literatürün taranması.
6. Hipotezlerin yazılması.
7. Program değerlendirmeyle ilgili araştırma için model ve yaklaşımların belirlenmesi.
8. Evren ve örneklem belirlenmesi.
9. Veri toplamak için tekniklerin belirlenmesi.
10. Verilerin toplanarak çözümlenmesi ve yorumlanması.
11. Değerlendirme araştırmasının tamamlanması, raporlaştırılması ve paydaşlara duyurulması.
12. Değerlendirmenin değerlendirilmesi (Meta Evaluation).

Program değerlendirme sürecinin son aşaması olan değerlendirilmenin değerlendirilmesinde amaçlanan değerlendirmede yapılan noksanlık ve yanlışlıkları belirlemektir. Çalışmadaki noksanlık ve yanlışlıkların belirlenmesi ile çalışmaların niteliği hakkında bilgi edinilmektedir. Bu aşamaya meta değerlendirme adı da verilmektedir (Fitzpatrick, Sanders ve Worthen, 2004). Meta Değerlendirme, bir değerlendirme sürecinin ve sonuçlarının nasıl savunulacağına dair soruları yanıtlamanın bir yoludur ve bu nedenle değerlendirme verilerinin kullanımını artırma potansiyeline sahiptir (Cooksy ve Cacarelli, 2005). Ayrıca sonuçların daha nicel verilere dayandırılmasını sağlar (Lynch, Greer, Larson,, Cummings, Harriett, Dreyfus, ve Clay, 2003). Değerlendirme çalışmaları, verilerin ölçüt ve standartlar ile karşılaştırılması sonucu yapılmaktadır. Bahsi geçen ölçüt ve standartlar genel olarak evet-hayır şeklinde kesin çizgiler ile birbirinden ayrılmış bulunmaktadır. Fakat, bu derecelendirme ile evet ile hayır arasındaki noktalar göz ardı edilmektedir. Değerlendirmelerin daha nesnel olması için bir kontrol mekanizmasına ihtiyaç duyulmaktadır. Bu mekanizmanın meta değerlendirme ile sağlanması mümkündür (Semerci, 2008). Meta değerlendirme kavramsallaştırılmasından önce, meta değerlendirme ile meta-analizi karşılaştırmak faydalı olabilmektedir. Bu terimler oldukça farklı kavramlara atıfta bulunsa da genellikle uygunsuz bir şekilde eşitlenmektedir. Meta değerlendirme, belirli bir değerlendirmenin değerini değerlendirir. Öte yandan, bir meta-analiz, ortak bir araştırma sorusunu ele alan çalışmaların nicel bir sentezi konumundadır (Stufflebeam, 2001, 187).

Michael Scriven, 1969'da "Educational Product Report" da meta değerlendirme kavramını tanıtmıştır. Bu kavramı, eğitim ürünlerini değerlendirmek için bir planın değerlendirmesine atıfta bulunmak amacıyla kullanmıştır. Esasen, Dr. Scriven meta değerlendirmeyi bir değerlendirme, değerlendirme sistemi veya değerlendirme cihazının herhangi bir değerlendirmesi olarak tanımlamıştır. Yanlış ve/veya taraflı raporların yayımlanmasının, tüketicileri değersiz veya kalitesiz eğitim ürünleri satın almaya iteceğini ve bu durumun çocukların ve gençlerin zararına olacak şekilde yanlış yönlendirmeye neden olacağını savunmuştur. Bu nedenle, bu tür ürünlerin değerlendirmelerinin değerlendirilmesi gerektiğini ve bu tür meta değerlendirmelerin tüketicilerin refahı için kritik öneme sahip olduğunu vurgulamıştır (Stufflebeam, 2001, 185). Scriven (1969) meta-değerlendirme kavramını "ikincil değerlendirme" veya diğer bir tanımlamayla "değerlendirmenin değerlendirilmesi" olarak ifade ederken meta-değerlendirmeyi kuramsal açıdan değerlendirmeyi yöntemsel değerlendirme olarak görürken uygulama bakımından özel performans değerlendirmesiyle alakalı olduğunu belirtmektedir (Cooksky ve Caraceli, 2005). Meta değerlendirme, bir değerlendirme faaliyetini tanımlamaya ve onu iyi değerlendirmenin nasıl olması gerektiğine ilişkin bir dizi düşünceye karşı yargılamaya yönelik bir prosedür

olarak tanımlanmıştır (Stufflebeam, 2011, 135). Yılmaz (2021, 1474) ise meta değerlendirmeyi, mevcut araştırmaların raporlaştırılması esnasında ortaya çıkan eksiklere dikkat çekilerek, araştırmaların kalitesinin artırılması için çalışmaların niteliğinin önceden belirlenen kriterlere göre değerlendirilmesinin yapıldığı çalışmalar olarak tanımlamıştır. Diğer bir ifadeyle ise Stufflebeam (1978) değerlendirmenin uygulanabilirlik, faydalılık, etik ve teknik yönden yeterli olabilmesiyle ilgili tasvir ve yargıyla alakalı bilgileri toplayarak kullanma süreci olarak meta değerlendirmeyi tanımlamaktadır. Buradan değerlendirme sonuçlarının etkili kullanılması ve değerlendirme süreciyle birlikte sonuçlarının alakalı olan çevreyle paylaşılması da önemli bir işlev olarak görülmektedir (Cooksky ve Caraceli, 2005). Bu noktada, değerlendirmenin toplumun vicdanı olması gibi, meta değerlendirmenin de değerlendirmenin vicdanı olduğu söylenebilmektedir (Scriven, 2009, 8).

Brinkerhoff, Brethower, Nowakowski ve Hluchyj (1983) meta-değerlendirmenin amaçlarını şu şekilde ifade etmektedirler:

- Değerlendirme amacı ve sorularının değerini ve güçlüğünü değerlendirmek.
- Değerlendirme deseninin güçlüğünü ve uygunluğunu belirlemek.
- Bilgi toplamaya yönelik yöntemlerle ulaşılan bilgilerin niteliği ve konuyla ilgili olma seviyesini değerlendirmek.
- Analizlerin ve analizlerle ilgili yorumların yeterlilik ve doğruluğunu değerlendirmek.
- Değerlendirmeye ilgili raporu düzen, faydalılık ve yeterlilik açısından değerlendirmek.
- Harcamaların onaylanabilirliğini belirlemek amacıyla değerlendirmenin verimli bir şekilde ilerleyip ilerlemediğini ve ekonomikliğini değerlendirmek.

Geçerli ve güvenilir niteliklere sahip yürütülecek meta-değerlendirme çalışmasının aşamaları aşağıda verilmiştir (Stufflebeam, 2000):

Değerlendirme paydaşlarının belirlenmesi: Meta-değerlendirmede değerlendirme neticelerine ihtiyaç duyanlar ve değerlendirme çalışmasına iştirak edenler paydaş olarak adlandırılmaktadır. Araştırmacılar, aileler, öğretmenler, öğrencilerle birlikte özel kurumlar vb. paydaşlar meta-değerlendirmede görülebilmektedir. Çalışmanın nitelikli olmasında paydaşların aşikâr olması beklendiğinden çalışma öncesinde paydaşların belirlenmesi gereklidir.

Değerlendiricilerin belirlenmesi: Değerlendirme sonuçlarını kullanacak ve değerlendirmeyi isteyen kurumun alanında yetkin değerlendirme uzmanlarını belirlediği aşamadır. Değerlendirme uzmanlarının farklı uzmanlık alanlarının bulunması (değerlendirme istatistiği, değerlendirme modelleri alan uzmanı vb.) çalışmanın nitelikli olmasını sağlamaktadır.

Meta-değerlendirme sorularının belirlenmesi: Değer ve faydalılık ölçütleri gözetilerek araştırılacak soruların belirlenmesidir. Bu iki ölçütle değerlendirmenin etkili olup olmadığı ve değerlendirme sonuçlarından yararlanacak paydaşların ihtiyaçlarının karşılanıp karşılanmadığıyla ilgili niteliğine bakılmaktadır.

İlgili araştırmaların toplanarak gözden geçirilmesi: İlgili araştırmaların belirlenmesi detaylı bir kaynak taramasını gerektirmektedir. Araştırmacıların geçerli ve güvenilir nitelikte sonuçlara ulaşabilmesi için bütün verilere erişebilmesi gereklidir. Bundan dolayı araştırmanın gerçekleştirilmesinde seçilecek amaç, yer, zaman ve katılımcılar önem arz etmektedir. Özetle değerlendirici geniş kapsayıcı bilgilere sahip olarak araştırmanın değerlendirme standartlarına uygun olmasına dikkat etmelidir.

Bulguların analiz edilmesi: Ayrıntılı taranan araştırma sonuçlarındaki bilgilerin elverişli analiz yöntemleri kullanılarak çözümlenmesi gereklidir. Burada verilerin analiz edilmesinde nicel ve nitel yöntemlerin her ikisi de kullanılabilir. Genellikle hedeflenen verilerin tekrar analizi, bilgisayar destekli konu analizi vb. teknikler tercih edilmekle birlikte pasta ve bar grafikleri vb. olmak üzere daha çok görseller tercih edilmektedir. Bununla değerlendirme çalışmasındaki verilerin bütün olarak toplanması ve program değerlendirme standartlarına uygun hale getirilmesi amaçlanmaktadır.

Bulguların standartlarla karşılaştırılması: Program değerlendirme standartları ve ulaşılan analiz sonuçlarının karşılaştırılması aşamasıdır. Bunun için farklı uzmanlar tarafından uyarlanmış ölçek ve anketlerden faydalanılabilmektedir. Bununla beraber hedeflenen program değerlendirme çalışmaları uygunluk, uygulanabilirlik, doğruluk ve faydalılık öğeleri açısından ölçülmektedir. Aşamaların neticelenmesiyle birlikte veriler yeniden analiz edilir. Tercih edilecek ölçek veya anketin yapısına bağlı olarak çalışmada tekil ya da ilişkisel analiz yöntemleri seçilebilmektedir.

Raporun hazırlanması: Verilerin analiz edilmesiyle beraber yapılan öneri ve yorumların kullanılması ve ilgililere duyurulması amacıyla çalışmanın bütün boyutlarını içine alan bir raporun hazırlandığı aşamadır.

Yağan (2019), program değerlendirme alanında yayınlanan dokuz doktora tezinin meta değerlendirilmesini gerçekleştirmiştir. Değerlendirmede Stufflebam'ın (2012), meta değerlendirme kontrol listesinden dört boyut, 26 standart ve 206 alt standart kullanıldığı görülmektedir. Çalışmanın sonuçları incelendiğinde değerlendirici güvenilirliği, değerlere saygı, süreç ve ürünler, proje yönetimi, yasal çerçeve, katılımcı hakları, geçerli bilgi ve bilgi yönetimi standartlarına yönelik içeriklere yer verildiği görülürken; paydaş odaklılık, amaçlar, verilerin niteliği, iletişim, sonuçlar ve etki, uygulama süreçleri, bağlamsal sürdürülebilirlik, kaynak kullanımı, oryantasyon, açıklık ve yalınlık, şeffaflık ve bilgilendirme, çıkar çatışması, sonuç ve kararlarda doğruluk, güvenilirlik, programı betimleme, değerlendirme deseni ve analiz, istatistiksel kanıtlar ile raporlama standartları açısından istenilen düzeyde yer verilmediği görülmektedir. Yağan'ın (2019), bu araştırmasının, ulusal düzeyde meta değerlendirme konusunda yapılmış az sayıda çalışmadan biri olduğu için ve aynı zamanda meta değerlendirme çalışmaları tamamlanmış değerlendirmelerin kalitesini de ortaya koyduğu için önemli olduğu ve literatüre katkı sağladığı düşünülmektedir.

Araştırmanın Amacı

Araştırmanın amacı, program değerlendirme tezlerinin Rasch ölçme modeli ile meta değerlendirmesinin yapılmasıdır. Bu amaç doğrultusunda;

1. Program değerlendirme tezlerinin kalibrasyon haritası,
2. Program değerlendirme tezlerinin analizi,
3. Jürilerin sertlik veya hoşgörülerine ilişkin analizi,
4. Program değerlendirme ölçütlerindeki maddelere ilişkin madde güçlük analizi ve,
5. Jürilerin tarafsızlık analizi yapılmıştır.

Yöntem

Araştırma Modeli

Araştırmada tarama modeli kullanılmıştır. Araştırmada ayrıca Linacre (1993) tarafından geliştirilen Çok Yüzeyle Rasch Ölçme Modeli (ÇYRÖM) kullanılmıştır.

Çalışma Grubu

Çalışma grubunu, 2020-2021 bahar akademik yılında Eğitim Programları ve Öğretim bilim dalında yüksek lisans ve doktora yapan toplam 5 öğrenci oluşturmaktadır.

Veri Toplama Araçları

Çalışmada veri toplama aracı olarak program değerlendirme ölçütleri kullanılmıştır. Program değerlendirme ölçütleri 12 kriterden oluşmaktadır. Bu ölçütler dokuz uzman görüşüyle belirlenmiştir. Jüriler ilgili ölçütleri "Hiç yeterli değil: 1", "Çok az yeterli: 2", "Kısmen yeterli: 3", "Büyük oranda yeterli: 4", "Tamamen yeterli: 5" derecelendirmeleriyle puanlamışlardır.

Verilerin Toplanması

Veriler 2020-2021 bahar akademik yılında Yükseköğretim Kurulu tez taramasında program değerlendirme anahtar kelimesiyle bulunan 16 tezden elde edilmiştir.

Verilerin Analizi

Program değerlendirme ile ilgili yapılan tezlerin analizi, çok-yüzeyle Rasch ölçme modelinde fazlaca kullanılmakta olan bir program olarak öne çıkan ve Linacre (1993, 2-15) tarafından geliştirilen FACETS analiz programı ile gerçekleştirilmiştir. Bu program kullanılırken Linacre (2008)'nin kılavuz kitabından yararlanılmıştır.

Bulgular

Program değerlendirme alanında yapılmış tezlerin, Rasch ölçme modeli ile analizi sonucu ortaya çıkan, çalışmaya ait yüzeyler (program değerlendirme alanında yapılmış tezler, jürilerin sertlik/hosgörü durumu ve

+ -2	+Yetersiz nitelik	+ Sert	+ Basit
Logit	+Tezler	+Jüri	-Maddeler

Şekil 1. Program değerlendirme tezlerinin kalibrasyon haritası

Şekil 1, kalibrasyon haritasında çalışmaya dair yüzeyle ait genel bir analizi göstermektedir. Şekil 1 incelendiğinde tez12 ve tez7'nin yüksek nitelikli tezler olduğu, tez13'ün ise düşük nitelikli tez olduğu söylenebilir. Çoğu tez çalışmalarının ise daha ortalama nitelikte çalışmalar olduğu söylenebilmektedir. Jürilerden en hoşgörülü jüri J4, en katı jüri ise J3'tür. J4 haricindeki jürilerin daha birbirine benzer bir puanlama davranışı gösterdiği söylenebilir. Kriterler açısından en zor kriterin "Değerlendirmede kullanılan program değerlendirme modelinin uygunluğu" olduğu görülmektedir. Kriterler açısından en basit kriterlerin ise "Değerlendirme sürecini tanımlama" ve "Değerlendirmede veri analizlerinin güvenilirliği" olduğu söylenebilir. Aşağıda, program değerlendirme tezlerinin ölçüm sonuçları gösterilmiştir.

Tablo 1. Program değerlendirme tezlerinin ölçüm raporu

Toplam Puan	Gözlem Sayısı	Gözlenen Ortalama	Yansız Ortalama	Model Ölçüm	Model Hata	Uygunluk İçİ MnSq	Uygunluk İçİ ZStd	Uygunluk Dışı MnSq	Uygunluk Dışı ZStd	Nu/tezler	
295	60	4.92	4.94	1.46	.42	1.26	.5	2.33	1.5	12 tez12	
293	60	4.88	4.92	1.17	.35	.73	-.3	1.35	.7	7 tez7	
282	60	4.70	4.81	.40	.21	1.08	.3	1.94	1.7	3 tez3	
279	60	4.65	4.78	.28	.19	1.30	.9	2.66	2.7	11 tez11	
277	60	4.62	4.77	.21	.19	1.13	.5	2.61	2.7	2 tez2	
276	60	4.60	4.76	.18	.18	1.05	.2	.81	-.3	1 tez1	
274	60	4.57	4.74	.11	.18	1.26	.9	1.87	1.8	5 tez5	
274	60	4.57	4.74	.11	.18	1.17	.6	1.31	.8	9 tez9	
271	60	4.52	4.71	.03	.17	1.28	1.0	.97	.0	6 tez6	
271	60	4.52	4.71	.03	.17	1.36	1.3	.86	-.2	8 tez8	
265	60	4.42	4.64	-.13	.15	1.28	1.1	1.84	1.9	10 tez10	
244	60	4.07	4.38	-.55	.13	.64	-2.0	.54	-1.7	16 tez16	
239	60	3.98	4.30	-.63	.13	.92	-.3	.86	-.3	15 tez15	
234	60	3.90	4.22	-.71	.12	.55	-2.7	.54	-1.8	14 tez14	
227	60	3.78	4.11	-.82	.12	.50	-3.3	.47	-2.4	4 tez4	
202	60	3.37	3.60	-1.17	.12	.72	-1.7	.65	-1.6	13 tez13	
Gözlenen Puan	Gözlem Sayısı	Gözlenen Ortalama	Yansız Ortalama	Model Ölçüm	Model Hata	Uygunluk İçİ MnSq	Uygunluk İçİ ZStd	Uygunluk Dışı MnSq	Uygunluk Dışı ZStd	No/Tezler	
262.7	60.0	4.38	4.57	.00	.19	1.01	-.2	1.35	.3	Ortalama (N: 16)	
25.9	.0	.43	.36	.69	.08	.29	1.5	.76	1.7	Standart sapma.	
RMSE (Model)	: .20	Ayırma indeksi:	3.22				Güvenirlilik:	.91			
Tamamı aynı ki-kare	: 184.5	Sd= 15	p= .00								
Normal ki-kare	: 12.9	Sd=14	p= .53								

Program Değerlendirme Tezlerinin Analizi

Tablo 1'de, program değerlendirme alanında yapılmış tezlerin incelenmesine ait detaylı bir ölçüm raporu sunulmuştur. Tablo incelendiğinde logit değerlerine ait standart hata (RMSE, Root Mean Square Standart Error)

değeri 0.20 ve standart sapma da kritik değer olan 1.0'ın altında görülmektedir. Rasch analizi sonucunda güvenilirlik katsayısı 0.91 olarak elde edilmiştir. Bulunan bu katsayı, hangi güvenilirlikle program değerlendirme alanındaki tezlerin sıralandığını belirtmektedir. Bu 0.91'lik katsayı, program değerlendirme alanındaki tezlerin oldukça yüksek bir güvenilirlikle sıralandığını göstermektedir. Ayırma indeksi 3.22 ve güvenilirlik katsayısı 0.91 ile sabit etkiye ait "Program değerlendirme alanında yapılan tezler arasında anlamlı bir fark vardır" hipotezi ki-kare ile test edildiğinde ($\chi^2=184.5$, $sd=15$, $p=0.00$) yokluk hipotezi reddedilmiştir. Böylece, program değerlendirme alanında yapılan tezler arasında istatistiksel açıdan anlamlı farklılıklar bulunduğu sonucuna ulaşılmıştır.

Rasch analizinde "Uygunluk içi" ve "Uygunluk dışı" değerleri için ortaya konulan kalite kontrol sınırı 0.6 ile 1.4 olarak belirlenmiştir (Wright and Linacre, 1994, 375-380). Karar vermek için beklenmeyen cevaplar "Uygunluk içi" ve uzaktaki beklenmeyen cevaplara hassaslık içeren "Uygunluk dışı" değerleri değerlendirmeye alınmaktadır (Baştürk ve Işıkoğlu, 2007, 737). Çalışma bu değerler ışığında incelendiğinde, uygunluk gösterdiği görülmektedir. Çalışma sonuçlarına göre, alınan toplam 16 tez için 10 tezin (tez12, tez7, tez3, tez11, tez2, tez1, tez5, tez9, tez6, tez8) yüksek nitelikli ve 6 tezin (tez10, tez16, tez15, tez14, tez4, tez13) düşük nitelikli olduğu söylenebilir (Şekil 1).

Jürilerin Sertlik/Hoşgörülerine İlişkin Analizi

Jürilerin sertlik/hoşgörü karşılaştırması Tablo 2'de verilmiştir. Tabloda jüri ayırma indeksi 6.18 ve güvenilirlik katsayısı .97 ile sabit etkiye ait "Jürilerin sertlik/hoşgörülerini arasında farklılık vardır" hipotezi ki-kare ile test edildiğinde ($\chi^2=66.7$ $sd=4$ $p=.00$) yokluk hipotezi reddedilmiştir. Bu durum, beş jürinin puanlamalarının sertlik/hoşgörülerini arasında istatistiksel olarak anlamlı farklılık olduğunu göstermektedir.

Tablo 2. Jürilerin sertlik/hoşgörü karşılaştırması

Toplam Puan	Gözlem Sayısı	Gözlenen Ortalama	Yansız Ortalama	Model Ölçüm	Model Hata	Uygunluk İçi MnSq	Uygunluk İçi ZStd	Uygunluk Dışı MnSq	Uygunluk Dışı ZStd	No/Jüri
938	75	4.64	4.66	1.75	.20	.87	-.4	.77	-.8	4 J4
829	75	4.64	4.66	1.75	.20	1.01	.1	.98	.0	5 J5
820	75	4.55	4.58	1.50	.18	.44	-2.8	.48	-2.50	1 J1
818	75	4.52	4.55	1.44	.17	1.20	.8	1.13	.6	2 J2
798	75	4.51	4.54	1.42	.17	.51	-2.3	.55	-2.2	3 J3
Gözlenen Puan	Gözlem Sayısı	Gözlenen Ortalama	Yansız Ortalama	Model Ölçüm	Model Hata	Uygunluk İçi MnSq	Uygunluk İçi ZStd	Uygunluk Dışı MnSq	Uygunluk Dışı ZStd	No/Jüri
840.6	192.0	4.38	4.62	1.52	.10	1.0	-.4	1.35	.2	Ortalama (N: 5)
49.7	.0	.26	.16	.64	.05	.22	1.2	.98	2.0	Standart Sapma
RMSE (Model)		: .11	Ayırma indeksi: 6.18		Güvenirlik: .97					
Tamamı aynı ki-kare		: 66.7	Sd= 4	p= .00						
Normal ki-kare		: 3.8	Sd= 3	p= .29						

Tablo 2 incelendiğinde, jürilerin en hoşgörülü olandan en sert olana doğru sıralandığında Jüri4'ün (J4) en hoşgörülü ve Jüri3'ün (J3) de en sert jüri olduğu görülmektedir. Jüri numaralarına göre, en hoşgörülünden en sert olana doğru J4> J5> J2> J1> J3 şeklinde sıralanabilir. Bu duruma ek olarak, Şekil 1 incelendiğinde, jürilerin logit değerlerinin sertlik/hoşgörü ölçeğinde 1 logit uzaklıkta kümelenmiş olması jürilerin sertlik/hoşgörü durumlarına dair farklarının epeyce az olduğunu göstermektedir (Lee ve Kantor, 2003 akt. Atılğan, 2005, 66). Elde edilen bu sonuç, beş jürinin 16 program değerlendirme alanındaki tezlere ilişkin yaptıkları puanlamalarının anlamlı düzeyde farklılık içerdiği, ancak göreceli olarak birbirine yakın puanlar verdiklerine işaret etmektedir.

Program Değerlendirme Ölçütlerine Ait Madde Güçlük Analizi

Tablo 3'te program değerlendirme ölçütlerine ilişkin madde güçlük analizine yönelik istatistikler sunulmuştur. Ayırma indeksi 3.51 ve güvenilirlik katsayısı .92 ile "program değerlendirme ölçütlerine ait maddelerin güçlükleri

arasında anlamlı bir fark vardır” hipotezi ki-kare ile test edildiğinde ($\chi^2=894.3$ $sd=24$ $p=.00$) yokluk hipotezi reddedilmiştir. Elde edilen durum gösteriyor ki, program değerlendirme ölçütlerine ait maddelerin güçlükleri arasında istatistiksel olarak anlamlı farklılıklar içermektedir.

Tablo 3. Meta değerlendirmede kullanılan ölçütlerin madde güçlük istatistikleri

Toplam Puan	Gözlem Sayısı	Gözlenen Ortalama	Yansız Ortalama	Model Ölçüm Hata	Uygunluk İçü MnSq	Uygunluk Dışı MnSq	No/Maddeler
286	80	3.58	3.92	.96 .10	1.28 1.7	2.16 3.9	3 Model uygundur
298	80	3.72	4.09	.83 .11	1.15 .9	1.18 .7	12 Ölçme aracı uygundur
323	80	4.04	4.39	.53 .11	.81 -1.1	.76 -.8	9 Araştırmacının eğitimi yeterli
341	80	4.26	4.56	.28 .12	.78 -1.2	.80 -.5	10 Bulgular karşılaştırılmıştır
354	80	4.43	4.67	.06 .14	.82 -.8	.84 -.3	7 Teknik bakımdan yeterlidir
355	80	4.44	4.68	.04 .14	.79 -1.0	.83 -.4	6 Literatüre katkı sağlamıştır
363	80	4.54	4.74	-.12 .15	.88 -.4	2.59 3.1	8 Kullanılan dil açık ve nettir
369	80	4.61	4.78	-.26 .16	.95 .0	.67 -.7	4 Ölçme araçları geçerlidir
371	80	4.64	4.79	-.31 .17	.74 -.9	4.01 4.4	2 Öncelikler belirlenmiştir
374	80	4.68	4.81	-.40 .17	.72 -.9	.60 -.9	11 Öneriler uyumludur
384	80	4.80	4.88	-.78 .22	.62 -1.0	.42 -1.3	5 Veri analizleri güvenilirdir
385	80	4.81	4.89	-.84 .23	.98 .0	1.32 .7	1 Süreç tanımlanmıştır

Gözlenen Puan	Gözlem Sayısı	Gözlenen Ortalama	Yansız Ortalama	Model Ölçüm Hata	Uygunluk İçü MnSq	Uygunluk Dışı MnSq	No/Maddeler
350.3	80.0	4.38	4.60	.00 .15	.87 -.4	1.35 .6	Ortalama (N: 12)
32.4	.0	.41	.31	.57 .04	.19 .9	1.06 2.0	Standart Sapma

RMSE (Model)	: .16	Ayrırma indeksi:	3.51	Güvenirlilik:	.92
Tamamı aynı ki-kare	: 171.7	Sd=	11	p=	.00
Normal ki-kare	: 10.2	Sd=	10	p=	.43

Tablo 3’de, jürilerin program değerlendirme alanında yapılan tezlerin meta değerlendirmesini gerçekleştirmede kullandıkları program değerlendirme ölçütlerinde en çok zorlandıkları maddeler, öncelikle “Değerlendirmede kullanılan program değerlendirme modeli uygundur.” olmuştur. Tablo 3’e göre jürilerin en basit uyguladıkları maddeler ise, “Değerlendirme süreci tanımlanmıştır.” ve “Değerlendirmede veri analizleri güvenilirdir.” olarak belirlenmiştir.

Jürilerin Tarafsızlık Analizi

Jürilerin tarafsızlık analizi Tablo 4’te verilmiştir. Tabloda yer alan Z puanlarının +2 ve -2 dışında bulunması bir yanlılığının göstergesidir. Tablo 4 incelendiğinde, Z puanlarının 1.72 ile -4.59 arasında değiştiği görülmektedir.

Tablo 4. Jüri ile program değerlendirme tezlerinin etkileşim analizi

Toplam Puan	Beklenen Puan	Gözlem Sayısı	Göz/Bek. Ortalama	Yanlılık+ Ölçme	Model Hata	Z	U. içü MnSq	U. dışı MnSq	Sq	Nu	Tezler	Ölç.	N	Jüri	Ölç.
57	59.7	12	-.23	-2.37	.52	-4.59	.6	.3	60	12	tez12	1.46	4	J4	2.79
40	53.2	12	-1.10	-1.02	.24	-4.17	.4	.4	6	6	tez6	.03	1	J1	1.23
57	59.2	12	-.19	-1.19	.52	-2.30	1.4	2.0	59	11	tez11	.28	4	J4	2.79
56	58.8	12	-.23	-1.00	.44	-2.27	1.1	1.6	58	10	tez10	-.13	4	J4	2.79
57	59.1	12	-.18	-1.11	.52	-2.16	1.3	2.2	50	2	tez2	.21	4	J4	2.79
35	43.5	12	-.71	-.51	.24	-2.12	.6	.6	47	15	tez15	-.63	3	J3	1.08
49	54.6	12	-.47	-.59	.28	-2.09	1.0	1.0	2	2	tez2	.21	1	J1	1.23
28	36.7	12	-.73	-.51	.25	-2.04	.4	0.4	29	13	tez13	-1.17	2	J2	1.21
59	59.7	12	-.06	-1.15	.95	-1.21	.9	1.2	55	7	tez7	1.17	4	J4	2.79
60	58.0	12	.16	1.58	1.69	.94	.0	.0	64	16	tez16	-.55	4	J4	2.79

Program Değerlendirme Tezlerinin Meta Değerlendirmesi

60	57.3	12	.22	1.70	1.55	1.10	.0	.0	52	4	tez4	-.82	4	J4	2.79
60	55.8	12	.35	2.25	.73	1.30	.0	.0	61	13	tez13	-1.17	4	J4	2.79
59	53.2	12	.48	1.57	.95	1.65	.9	.1	8	8	tez8	.03	1	J1	1.23
58	51.9	12	.50	1.11	.64	1.72	.8	.7	38	6	tez6	.03	3	J3	1.08

Gözlenen Puan	Beklenen Puan	Gözlem Sayısı	Göz/Bek. Ortalama	Yanlılık+ Ölçme	Model Hata	Z	U.İçi MnSq	U.dışı MnSq	Tezler Ölç. N	Jüri Ölç.
52.5	52.56	12.0	.00	.10	.56	-.10	0.8	0.9	Ortalama (N:80)	
7.4	6.28	.0	.29	.69	.39	1.26	0.5	0.7	Standart Sapma	

Tamamı aynı ki-kare: 125.4 Sd= 80 p= .00

Tarafsızlık analizine bakıldığında, jürilerden bazılarının program değerlendirme alanında yapılan tezlerden bazılarına karşı oldukça sert veya hoşgörülü puanlama yaptıkları görülmektedir. Program değerlendirme alanında yapılan tezlerin değerlendirmesinde 12 numaralı tez12'nin 4 numaralı jüriden (J4) yaklaşık 60 puan alması beklenirken 57 puan alması ile jürinin sert puanlama yaptığını tespit edilmiştir (Z=-4.59). 6 numaralı tez6'nın 1 numaralı jüri (J1) tarafından yaklaşık 53 puan alması beklenirken 40 puan alması (Z=-4.17), yine 11 numaralı tez11'in 4 numaralı jüri (J4) tarafından yaklaşık 59 puan alması beklenirken 57 puan alması (Z=-2.30) bu jürilerin sert puanlama yaptığını göstermektedir. Benzer şekilde program değerlendirme alanında yapılan tezlerin değerlendirmesinde 6 numaralı tez6'nın 3 numaralı jüri (J3) tarafından yaklaşık 52 puan alması beklenirken 58 puan alması (Z=1.72), 8 numaralı tez8'in 1 numaralı jüri (J1) tarafından yaklaşık 53 puan alması beklenirken 59 puan alması (Z=1.65), 13 numaralı tez13'ün 4 numaralı jüri tarafından (J4) yaklaşık 56 puan alması gerekirken 60 puan alması (Z=1.300), bu jürilerin oldukça hoşgörülü puanlama yaptığını göstermektedir. Bu taraflı davranışların sebepleri farklılık arz edebilir. Rasch ölçme modeli, hangi jürinin hangi ürüne karşı taraflı davrandığını göstermektedir. Fakat, tarafsızlığın nedenlerini bulmak için araştırmacıların ayrıca çalışmalar yapması gerekmektedir.

Sonuç, Tartışma ve Öneriler

Bu çalışmada, çok yüzeyli Rasch ölçme modeli ile program değerlendirme tezlerinin meta değerlendirmelerinin analizi yapılmıştır. Araştırmalarda eş zamanlı olarak kullanılan yüzeyler (program değerlendirme alanında yapılmış tezler, jürilerin sertlik/hosgörü ve kriterlerin zorluk derecesi) ilgili bağlamlarda düzenlenmiştir. Çalışma sonuçlarına göre 12 numaralı tez12 en yüksek nitelikli, 13 numaralı tez13 ise en düşük nitelikli tez olarak bulunmuştur. Tez12 295 puan, tez13 202 puan almıştır. En hoşgörülü jüri J4 ve en sert jüri J3'tür. Program değerlendirme alanında yapılmış tezleri değerlendiren jürilerin cevapladığı en basit kriter "Değerlendirme süreci tanımlanmıştır (385 puan ve 4.89 tarafsızlık ortalaması)" ve "Değerlendirmede veri analizleri güvenilirdir (384 puan ve 3.98 tarafsızlık ortalaması)" kriterleri olmuştur. Değerlendirmecilerin cevapladığı en zor kriter ise "Değerlendirmede kullanılan program değerlendirme modeli uygundur (286 puan ve 3.92 tarafsızlık ortalaması)" kriteri olmuştur. Program değerlendirme alanında yapılan tezler; "Değerlendirmede veri analizleri güvenilirdir", "Değerlendirme süreci tanımlanmıştır", "Öneriler içerik ve çalışmanın sonuçlarıyla uyumludur", "Değerlendirmede öncelikler belirlenmiştir.", "Değerlendirmede ölçme araçları geçerlidir" gibi kriterleri yüksek oranda karşılamaktadır. Fakat "Değerlendirmede kullanılan program değerlendirme modeli uygundur", "Ölçme aracının maddeleri kullanılan değerlendirme modeli ile uyumludur", "Araştırmacının program değerlendirme konusundaki eğitimi yeterlidir", "Elde edilen bulgular diğer akademik çalışma sonuçlarıyla karşılaştırılmıştır", "Değerlendirme teknik bakımından yeterlidir" gibi kriterleri ise istenilen oranda karşılamamıştır. Yükseköğretim Kurulu tez taramasında program değerlendirme anahtar kelimesiyle bulunan 16 tezin genel olarak nitelikli olduğu görülmektedir. Araştırmada Eğitim Programları ve Öğretim bilim dalında yüksek lisans ve doktora yapan toplam 5 öğrenci olan jürilerden bazıları tezleri değerlendirmede hoşgörülü davranmış, bazıları ise tez değerlendirirken sert davranmıştır. Jüri değerlendirmeleri incelendiğinde jürilerin sertlik/hosgörü farklılıklarının epeyce ılımlı olduğunu görülmektedir (Lee ve Kantor, 2003 akt. Atılğan, 2005, 66). Beş jürinin 16 program değerlendirme alanındaki tezler ile ilgili puanlamalarının anlamlı düzeyde farklı olduğu, buna karşın birbirine yakın puanlar verdikleri görülmüştür.

Araştırmaya dönük ve araştırmacılara yönelik olarak verilecek öneriler şunlardır;

1. Değerlendirmecilerin cevapladığı en zor kriter “Değerlendirmede kullanılan program değerlendirme modeli uygundur” olduğu düşünüldüğünde program değerlendirme alanında tez yazacak araştırmacıların program değerlendirme modelini araştırmalarına göre seçmeleri ve model hakkında detaylı araştırma yapmaları önerilmektedir.
2. Program değerlendirme alanında tez yazacak araştırmacıların kullanılan ölçme araçlarını seçerken daha titiz davranmaları, modele uygun olup olmadığını incelemeleri, modele uygun ölçme aracı geliştirmeleri önerilmektedir.
3. Tarafsızlık nedenlerinin ortaya çıkarılmasında çok yönlü Rasch ölçüm modeline ek olarak ölçme araçları kullanılması önerilmektedir.
4. Araştırmacının program değerlendirme konusundaki eğitimi büyük öneme sahip olduğundan, program değerlendirme alanında tez yazmak isteyenlere bu konuda eğitim alması ya da program değerlendirme alanında eğitim almış bir kişiye danışılması önerilmektedir.
5. Araştırmacılara elde edilen bulguların benzer akademik çalışma sonuçlarıyla karşılaştırılması, benzer çalışmalarla araştırmaların desteklenmesi önerilmektedir.

Kaynakça

- Atılğan, H. (2005). Müzik öğretmenliği özel yetenek seçme sınavının çok-yüzeyle Rasch modeli ile analizi (inönü üniversitesi örneği). *Eurasian Journal of Educational Research*, 0(20), 62- 73.
- Baştürk, R., Işıkoğlu Erdoğan, N. (2007). Okul öncesi eğitim kurumlarının işlevsel kalitelerinin çok- yüzeyle Rasch ölçme modeli ile analizi. *Kuram ve Uygulamada Eğitim Bilimleri*, 7(3), 727- 752.
- Brinkerhoff, R. O., Brethower, D. M., Nowakowski, J. ve Hluchyj, T. (Eds.). (1983). *Program evaluation: A practitioner's guide for trainers and educators* (Vol. 2). Springer Science & Business Media.
- Cooksky, L. ve Caraceli, V. (2005). Quality, context, and use: Issues in achieving the goals of metaevaluation. *American Journal of Evaluation*. 26 (1), 31–42.
- Erden, M. (1998). *Eğitimde program değerlendirme*. Ankara: Anı Yayıncılık.
- Ertürk, S. (2017). *Eğitimde program geliştirme* (2. baskı). Ankara: Edge Akademi.
- Fitzpatrick, J.L., Sanders, J.R ve Worthen, B. L. (2004). *Program evaluation alternative approaches and practical guidelines*. Boston, Pearson Education Inc.
- Demirel, Ö. (2020). *Eğitimde program geliştirme* (28. baskı). Ankara: Pegem Akademi Yayıncılık.
- Kaya, Z. (1997). Eğitimde program değerlendirme sürecinin temel işlemleri. *Gazi Üniversitesi, Endüstriyel Sanatlar Eğitim Dergisi*, 5(5), 59-72.
- Linacre, J. M. (1993). Generalizability theory and many facet Rasch measurement. *Annual Meeting Of The American Educational Research Association*. (April, 13, 1993), (ED 364 573). Atlanta Georgia.
- Linacre, J. M. (2008). *A user's guide to winsteps, ministep Rasch-model computer programs, program manuel*, 3.66, P.O. Box. 811322, Chicago IL 60681-1322.
- Lynch, D.C., Greer, A.G., Larson, L.C., Cummings, D.M., Harriett, B.S., Dreyfus, K.S. ve Clay, M.C. (2003). Descriptive metaevaluation. *Evaluation & The Health Professions*, 26(4), 447-461. DOI: 10.1177/0163278703258099
- Scriven, M. (1969). An Introduction to meta-evaluation. *Educational Product Report*, 2 (5), 36–38
- Scriven, M. (2009). Meta-evaluation revisited. *Journal of Multi Disciplinary Evaluation*, 6(11), iii-viii.
- Semerci, Ç. (2008). Meta değerlendirmeye farklı bir bakış: fazi düşünce. 17. ulusal eğitim bilimleri kongresi 01-03 Eylül, Sakarya Üniversitesi Eğitim Fakültesi, Sakarya/ Türkiye
- Stufflebeam, D. (1978). Meta-evaluation: an overview. *Evaluation and Health Profession*. 1 (1), 17–43.
- Stufflebeam, D. (2000). The methodology of metaevaluation as reflected in metaevaluations by the western michigan university evaluation center. *Journal of Personnel Evaluation in Education*, 14(1), 95–125.

- Stufflebeam, D.L. (2001). The metaevaluation imperative. *American Journal of Evaluation*, 22(2), 183–209.
- Stufflebeam, D. L. (2011). Meta evaluation. *Journal of Multi-Disciplinary Evaluation*, 7(15), 99-158.
- Stufflebeam, D. L. (2012). Program evaluations meta evaluation checklist. [Standards]. Erişim adresi:https://wmich.edu/sites/default/files/attachments/u350/2014/program_metaeval_short.pdf
- Uşun, S. (2016). *Eğitimde program değerlendirme süreçler yaklaşımlar ve modeller* (2. baskı). Ankara: Anı Yayıncılık.
- Wright, B.D., Linacre, J.M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370.
- Yağan, S. A. (2019). Program değerlendirme alanında yayınlanmış doktora tezlerinin meta değerlendirilmesi (2015-2018). *Electronic Journal of Education Sciences*, 8(16), 188-208.
- Yılmaz, K. (2021). Sosyal bilimlerde ve eğitim bilimlerinde sistematik derleme, meta değerlendirme ve bibliyometrik analizler. *Manas Sosyal Araştırmalar Dergisi*, 10(2), 1457-1490.

Extended Abstract

Curriculum evaluation (Erden, 1998), which is carried out to enable the development of the education program, is a process that allows to reach a decision about the effectiveness of the program as the last and complementary step (Ertürk, 2017) of the curriculum development process (Demirel, 2020). Program evaluation, as a result of this process, program development experts; It gives information about continuing the program, re-examining the program or moving to a new stage (Demirel, 2020). When program evaluation studies are carried out in accordance with a certain systematic, they allow the evaluation to be structured and made more comprehensive (Kaya, 1997). While evaluating a training program, a systematized scientific research process is carried out and this process can be handled in twelve basic stages (Uşun, 2016):

1. Recognizing the problem associated with program evaluation and identifying the issue of evaluation research.
2. Determining the purpose and research problem related to program evaluation.
3. Determining the time, possibilities and limitations related to the evaluation research.
4. Identification of internal and external evaluators.
5. Scanning the relevant literature.
6. Writing hypotheses.
7. Identifying models and approaches for research on program evaluation.
8. Determination of population and sample.
9. Identifying techniques for data collection.
10. Data collection, analysis and interpretation.
11. Completion, reporting and announcement of the evaluation research to the stakeholders.
12. Evaluation of evaluation (Meta Evaluation).

The aim of the evaluation, which is the last stage of the program evaluation process, is to bring the deficiencies and mistakes experienced in the evaluation to the surface. By determining the deficiencies and inaccuracies in the study, information about the nature of the studies is obtained. This stage is also called meta-evaluation (Fitzpatrick, Sanders, & Worthen, 2004). Meta-Assessment is a way of answering questions about how to defend an evaluation process and its results and therefore has the potential to increase the use of evaluation data (Cooksy & Cacarelli, 2005). It also allows the results to be based on more quantitative data (Lynch, Greer, Larson,, Cummings, Harriett, Dreyfus, & Clay, 2003).

Yağan (2019) carried out a meta-evaluation of nine doctoral theses published in the field of program evaluation. It is seen that four dimensions, 26 standards and 206 sub-standards from the meta-evaluation checklist of Stufflebam (2012) were used in the evaluation. When the results of the study are examined, it is seen that content related to evaluator reliability, respect for values, processes and products, project management, legal framework, participant rights, valid information and information management standards; stakeholder focus, objectives, quality of data, communication, results and impact, implementation processes, contextual sustainability, resource use, orientation, openness and simplicity, transparency and information, conflict of interest, accuracy in results and decisions, reliability, program description, evaluation pattern and analysis It is seen that statistical evidence and reporting standards are not included at the desired level. It is thought that this study by Yağan (2019) is important and contributes to the literature, as it is one of the few studies on meta-evaluation at the national level, and also because it reveals the quality of the evaluations whose meta-evaluation studies have been completed.

The aim of the research is to meta-evaluate program evaluation theses with the Rasch measurement model. In accordance with this purpose;

1. Calibration map of program evaluation theses,
2. Analysis of program evaluation theses,
3. Analysis of the juries on their harshness or tolerance,
4. Item difficulty analysis regarding the items in the program evaluation criteria and,
5. The impartiality analysis of the juries were made.

According to the results of the study, thesis 12 was found to be the highest qualified thesis, and thesis 13 was found to be the lowest qualified thesis. Thesis12 received 295 points and thesis13 received 202 points. The most generous rater is P4 and the most stringent rater is P3. The easiest criteria answered by the raters who evaluated

the theses made in the field of program evaluation were "Evaluation process is defined (385 points and 4.89 fairness average)" and "Data analysis in evaluation is reliable (384 points and 3.98 fairness average)" criteria. The most difficult criterion answered by the evaluators was the criterion "The program evaluation model used in the evaluation is appropriate (286 points and 3.92 fairness average)". It is seen that 16 theses are generally qualified. In the study, some of the raters, who were 5 graduate and doctorate students in Curriculum and Instruction, were generous in evaluating the theses, while others were strict while evaluating the thesis. When the raters' evaluations are examined, it is seen that the strictness/generosity differences of the raters are quite moderate (Lee and Kantor, 2003 cited in Atılgan, 2005, 66). The five raters' scores on the theses in 16 curriculum evaluation areas were significantly different, but they gave relatively close ratings.

Suggestions for research and researchers are as follows;

1. Considering that the most difficult criterion answered by the evaluators is "The curriculum evaluation model used in the evaluation is appropriate", it is recommended that researchers who will write a thesis in the field of curriculum evaluation should choose the curriculum evaluation model according to their research and conduct detailed research on the model.
2. It is recommended that researchers who will write a thesis in the field of curriculum evaluation should be more careful when choosing the measurement tools used, examine whether they are suitable for the model, and develop measurement tools suitable for the model.
3. It is recommended to use measurement tools in addition to the versatile Rasch measurement model to reveal the causes of bias.
4. Since the researcher's education on program evaluation is of great importance, it is recommended that those who want to write a thesis in the field of program evaluation should receive training on this subject or consult a program evaluation specialist.
5. It is recommended to compare the findings obtained from the research with the results of similar academic studies and to support the researches with similar studies.