

Köse, İ. A., Usta, H. G., Yandı A. (2016). Sunum yapma becerilerinin çok yüzeyle rasch analizi ile değerlendirilmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 16(4), 1853-1864.

Geliş Tarihi: 24/09/2016

Kabul Tarihi: 17/11/2016

## SUNUM YAPMA BECERİLERİNİN ÇOK YÜZEYLİ RASCH ANALİZİ İLE DEĞERLENDİRİLMESİ\*

İbrahim Alper KÖSE\*\*

Hatice Gonca USTA\*\*\*

Alperen YANDI\*\*\*\*

### ÖZ

Bu araştırmanın amacı öğrenci sunumlarına ait akran değerlendirmeleri kullanılarak puanlayıcı, ölçüt ve sunum becerilerinin çok yüzeyle Rasch modeli ile analizini yapmaktır. 2015-2016 eğitim-öğretim yılında eğitimde ölçme ve değerlendirme yüksek lisans dersini alan 6 öğrenci araştırma grubunu oluşturmaktadır. Öğrenciler sunum becerilerini kendileri ve akranları tarafından, araştırmacının puanlama kriterlerini belirlediği değerlendirme formunu kullanarak yapmışlardır. Analizler sonucunda puanlayıcıların katılık ve cömertlikleri bakımından ayrıldıkları, öğrencilerin konuyu sunmada ilgi çekici materyal kullanmadıkları, verilen sürele bütün katılımcıların uydukları belirlenmiştir. Sonuç olarak çok yüzeyle Rasch ölçme modelinin öğrenci performansını belirlemede kullanılabilir alternatif ölçme modeli olduğu örneklenmiştir.

**Anahtar kelimeler:** Çok yüzeyle Rasch ölçme modeli, sunum becerileri, madde tepki kuramı

## EVALUATION OF PRESENTATION SKILLS BY USING MANY FACETS RASCH MODEL

### ABSTRACT

The purpose of this study is to analyze the presentation skills of students, criterions and rater severity/leniency by using peer assessment by the help of many facets Rasch measurement model. 6 MA students were participated as rater and presenter to the study during the fall semester of 2015-2016 school years in the Measurement and Evaluation in Education course. Students are evaluated themselves and their friends using "Presentation Skills Evaluation Form". The results of the study demonstrated that raters are separated in terms of severity and leniency. Also students did not use interesting materials in their presentations but they obeyed time restrictions. As a result, many facet Rasch measurement model is an alternative model for evaluating student performances.

**Keywords:** Many facets Rasch measurement model, presentation skills, item response theory

---

\* Bu çalışma III nd International Eurasian Educational Research Congress (EJER-2016)'da bildiri olarak sunulmuştur.

\*\* Doç. Dr., Abant İzzet Baysal Üniversitesi, Eğitim Fakültesi, i.alper.kose@gmail.com

\*\*\* Yrd. Doç. Dr., Cumhuriyet Üniversitesi, Eğitim Fakültesi, goncausta@gmail.com

\*\*\*\* Araş. Gör., Abant İzzet Baysal Üniversitesi, Eğitim Fakültesi, alperenyandi@gmail.com

## 1.GİRİŞ

Okulda edinilen becerilerin günlük hayatta kullanılması gereksinimi ile öğrencilerde üst düzey davranışların kazandırılması ve bu davranışların ölçülmesi odak olmaya başlamıştır. Bilgilerin zihinde izlediği süreç bilgi, beceri ve yetenek olarak sıralandığında (Popham, 2008) klasik anlamda kâğıt kalem testleri beceri ve yetenek boyutundaki bu davranışları ölçmekte yetersiz kalmakta, performansa dayalı ölçme ve değerlendirme anlayışı ön plana çıkmıştır. Performansa dayalı ölçme ve değerlendirme araştırma yapma, sunum yapma, problem çözme, etkili iletişim kurabilme gibi üst düzey becerilerin gözlenmesine olanak sağlayan görevleri içermektedir. Bu üst düzey becerilerden biri de yükseköğretimde ve özellikle lisansüstü eğitimde çok sık kullanılan becerilerden olan sunum yapma becerisidir. Bireyin belirli bir konu hakkında dinleyicileri bilgilendirmeyi amaçlayan bu beceriler için çeşitli yazılımlar kullanılmaktadır. Gold Disk Astaund, Adobe Persuasion, Deltagraph Pro ve Microsoft PowerPoint (Baştürk, 2008) ve Prezi bunlardan bazılarıdır. Bu çalışmada yaygın kullanımından dolayı Microsoft PowerPoint merkezli sunumlar değerlendirilmiştir.

Bu görevlerin içeriğinde öğrenciler yapılandırılmış bir soru formatı ile karşılaşmadıkları için puanlama yapılırken objektif olunması da oldukça önemlidir. Bu bağlamda üst düzey becerilerin puanlanmasında kullanılan en önemli araç dereceli puanlama anahtarıdır. Puanlayıcılar, puanlama anahtarlarındaki performans kriterlerine göre dereceli olarak puanlamalarını yaparlar. Ancak puanlayıcının kararı sadece performansa dayalı değildir. Farklı faktörler/yüzeyler puanlamaya etki etmektedir. Bunlar; görevin/performansın zorluk derecesi, puanlayıcı katılığı/cömertliği, puanlananın geçmiş vb'dir. Ölçme konusu olmayan bu faktörlerin puanlamanın geçerliğini tehdit etmesi de kaçınılmazdır (Prieto ve Nieto, 2014).

Öğrenci becerilerini puanlarken puanlara çeşitli varyans kaynakları karışabilmektedir. Bu kaynaklar;

- 1- Puanlayıcı etkisi; katılık/cömertlik, halo etkisi, uç puanlardan kaçınma,
- 14- Performans için seçilen kriterlerin güçlüğü,
- 2- Puanlama kriterleri (Dereceleme ölçeğindeki noktaların puanlayıcılar arasında farklı anlamlar ifade etmesi).

Puanlayıcı etkisi puanlanan bireyin performansı ile ilgili olmayıp, puanlayıcı karakteristik özelliklerinden birisidir. Bu varyans kaynağı, ölçülen yapı ile ilgisi olmayıp, ölçümlere hata olarak karışan ve ölçme sonuçlarının geçerliğini tehdit eden önemli bir faktördür (Eckes, 2005). Çok yüzeyle Rasch modeli (ÇYRM) ile puanlayıcı karakteristikleri ölçme modeline yüzey (facet) olarak analize sokulabilmektedir.

### Çok Yüzeyle Rasch Modeli

Alanyazında bireylerin bir maddeye tepki vermelerinin altında yatan örtük özelliği açıklamak için iki temel kuramdan bahsedilmektedir. Bu kuramlar klasik test kuramı (KTK) ve madde tepki kuramıdır (MTK). Klasik test kuramı alanda uzun yıllardır kullanılmasına rağmen, karşılanması kolay zayıf varsayımlarının yanında, madde ve yetenek parametrelerinin gruba bağımlı olması, kestirimlerde herkes için tek bir hata kestiriminde bulunması gibi önemli sınırlılıkları da barındırmaktadır. Bu sınırlılıklara karşı geliştirdiği güçlü varsayımlarla son zamanlarda alanda sıklıkla kullanılan ölçme kuramlarından biri de Madde tepki kuramı olmuştur. Madde tepki kuramı bireyin

maddedeki performansı altındaki örtük özelliği matematiksel modeller yardımı ile açıklayan ölçme kuramıdır. Madde tepki kuramı modelleri çeşitli özelliklere göre sınıflandırılmaktadır; parametre sayısına göre, boyut sayısına göre, puanlama şekline göre (1-0, çoklu puanlanan...)...vs.

Çok yüzeyli Rasch modeli, temel Rasch modelinin bir uzantısıdır. Bunun iki nedeni vardır. Bunlar; Rasch modeli sadece iki yüzeye (cevaplayıcı ve maddeler) sınırlı olmasına rağmen, ÇYRM'de yüzey sınırlaması yoktur. İkinci olarak Rasch modeline ikili (1-0) puanlama söz konusuysa, ÇYRM çoklu puanlamaya izin verebilmektedir (Eckes, 2005). Rasch modeli, puanlama anahtarında yer alan seçenekler arası gerçek uzaklıkları hesaplamakta, daha hassas ve gerçekçi aralık birimi oluşturmaktadır (Elhan ve Atakurt, 2005). Bu anlamda Rasch modelinin objektiflik üzerine kurulduğu belirtilmektedir (Linacre, 1995).

Çok yüzeyli Rasch ölçme modelinde yer alan ve sistematik olarak test puanlarını etkileyen yüzey ifadesi faktör değişken veya bileşen olarak da tanımlanabilen ifadedir (Linacre, 2002).

#### 4.2. Araştırmanın Amacı

Tüm bunlar göz önüne alınarak araştırmanın amacını sunum becerilerinin çok yüzeyli Rasch modeliyle analiz edilmesi oluşturmaktadır. Bu genel amaç çerçevesinde;

- 1- Öğrencilerin sunum becerileri hakkında analizi,
- 2- Puanlayıcıların katılığı/cömertliği analizi,
- 3- Sunum becerilerine ilişkin görev güçlük analizi yapılmıştır.

#### 1.2. Araştırmanın Önemi

Yapılan bu çalışma ile birlikte öğrencilerin üst düzey düşünme becerilerinin objektif puanlamasına yönelik olarak eğitimcilerle destek sağlanacağı düşünülmektedir. Sunum yapma, yükseköğretim sürecinde sıkça kullanılmaktadır. Bu beceri ile ilgili olarak yapılan puanlamalara dayalı olarak öğrenciler ile ilgili kararlar alınmakta ve ders süreçlerinde düzenlemelere gidilmektedir. Alınan kararlar ve yapılan düzenlemelerin doğruluğu, yapılan ölçme işleminin geçerliği ile doğrudan bağlantılıdır. Geçerlik kavramı temelde ölçme sonuçlarının doğruluğu ile ilişkilidir. Ölçme sonuçlarının doğruluğunu etkileyen her olumsuz koşul geçerliği de olumsuz etkilemektedir. Geçerliği etkileyen olumsuz koşullardan biride puanlama sürecinde yapılan yanlışlık hatasıdır. Yapılan bu çalışma ile beraber puanlama sürecinde ortaya çıkabilecek hataların en aza indirilmesi hususunda eğitimcilerle bir öneri getirmek amaçlanmaktadır. Ayrıca eğitim sürecinde üst düzey düşünme becerilerinin daha az hata içerecek şekilde puanlanabilmesinin nasıl yapılabileceğine ilişkin alternatif bir yöntemle ilişkin bilgi verilmeye çalışılacaktır. Bu bağlamda yapılan çalışmanın gerek alanyazına gerek eğitimcilerle katkıları olacağından önemli olduğu düşünülmektedir.

## 2. YÖNTEM

### 2.1.Çalışma Grubu

Bu araştırmanın çalışma grubunu Abant İzzet Baysal Üniversitesi Eğitim Bilimleri Enstitüsü Ölçme ve Değerlendirme Yüksek Lisans Bilim Dalında 2015-2016 Eğitim

Öğretim yılında “Ölçme ve Değerlendirme” dersini alan 6 öğrenci ve bir öğretim elemanı oluşturmaktadır.

## 2.2.Araştırma Verileri

Öğrencilerin sunum becerilerini puanlamak için dersin çıktılarına paralel olarak araştırmacı tarafından geliştirilen “Sunum Becerileri Akran Değerlendirme Formu” kullanılmıştır. Araştırmaya 2015-2016 eğitim-öğretim yılında lisansüstü düzeyde “Ölçme ve Değerlendirme” dersini alan 6 öğrenci dâhil edilmiştir.

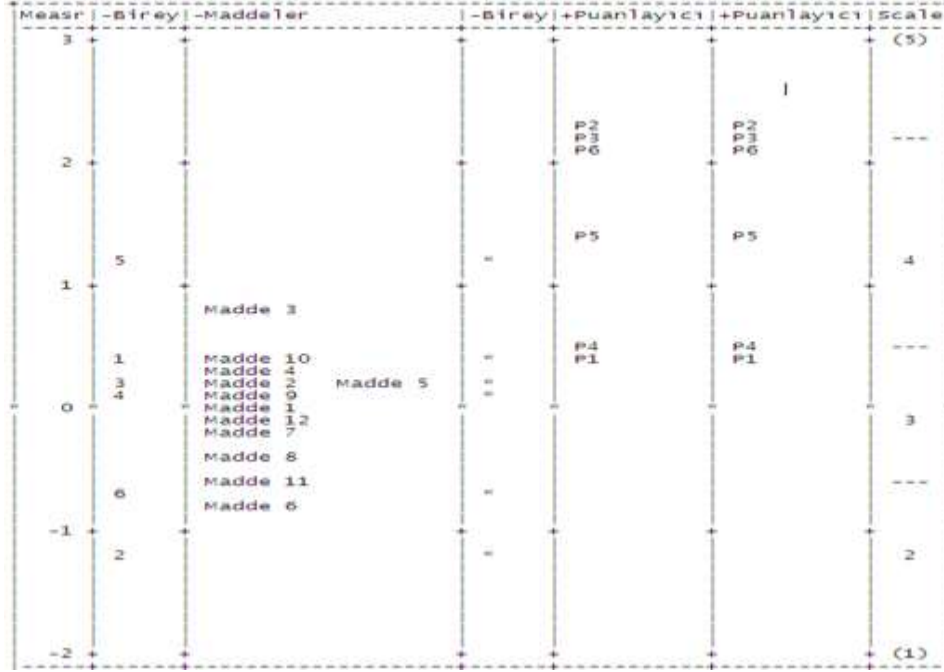
## 2.3.Verilerin Analizi

Araştırmada elde edilen gözlem verilerinin analizi için Linacre (1993) tarafından geliştirilmiş FACETS yazılımı kullanılmıştır. FACETS çok yüzeyle Rasch modeli için kullanılan yetenek, görev ve puanlayıcı gibi üç yüzeyi ele alan yaygın kullanımlı bir yazılımdır. FACETS doğru/yanlış, her zaman/bazen/asla, kesinlikle katılıyorum/katılıyorum/katılmıyorum gibi nicel puanları, saymaları, karşılaştırmaları, gözlemleri analiz eden bir yazılımdır. Her bir gözlem yüzeyler arasındaki ilişkinin bir sonucudur. FACETS yazılı performansın puanlanması, portfolyo değerlendirme ve puanlayıcıların puanladığı diğer performanslar için uygundur. Sadece eğitimde değil, ürün geliştirme, spor bilimleri ve sanat gibi farklı alanlar için de uygundur (Linacre, 1998). Bu araştırma kapsamında Rasch ölçme modelinde analizi yapılacak üç yüzey bulunmaktadır. Bunlar; puanlayıcı katılığı/cömertliği, sunum becerilerinde kullanılan 12 ölçüt ve 7 puanlayıcıdır. Değerlendirme formuna tepkiler “çok yetersiz” ile “çok yeterli” arasında değişen 5’li Likert düzeyinde verilmektedir.

## 3. BULGULAR

Sunum becerilerinin çok yüzeyle Rasch ölçme modeli ile analiz sonuçları incelendiğinde, araştırmada kullanılan yüzeyler (sunum performansı, puanlayıcı katılığı/cömertliği ve kullanılan ölçütler) ve bu yüzeylere ait genel bilgiler Şekil 1.’deki veri kalibrasyon haritasında verilmiştir. Şekil 1. incelendiğinde sunum yapan bireylerin, becerilerin değerlendirilmesinde kullanılan kriterlerin ve puanlayıcıların aynı logit cetvel üzerinde sıralandıkları görülmektedir.

Sunum yapan bireyler başarılarına göre en başarılıdan en başarısıza, becerilerin değerlendirilmesinde kullanılan maddeler en zordan en kolaya, puanlayıcılar ise en cömert üstte olmak üzere sıralanmışlardır.



Şekil 1. Veri Kalibrasyon Haritası

### 3.1. Lisansüstü Öğrencilerinin Sunum Becerilerinin Analizi

Tablo 1.'de 2015-2016 eğitim-öğretim yılında eğitimde ölçme ve değerlendirme yüksek lisans dersini alan 6 öğrencinin detaylı ölçüm raporu görülmektedir. Tablo incelendiğinde B5 kodlu birey sunum becerileri bakımından en başarılı, B2 kodlu birey ise en başarısız olarak değerlendirilmiştir. Öğrencilerin sunum becerilerine ilişkin logit değerlerine ait standart hata (RMSE) değerinin 0.17 olduğu görülmektedir. RMSE değeri aşırı uçlarda yer alan değerler hariç bütün değerler için ölçme hatasını verir (Baştürk, 2010). Hesaplanan RMSE değerinin düşük olduğu, düzeltilmiş standart hata değeri (0.77) 1'in altında hesaplanarak desteklenmiştir.

**Tablo 1. Lisansüstü Öğrencilerin Sunum Becerileri Analizi**

Total Score	Total Count	Obsvd Average	Fair(M) Average	Model Measure	S.E.	Infit MnSq	Zstd	Outfit MnSq	Zstd	Estim. Discrm	Correlation PtMea	PtExp	N Birey
232	72	3.22	3.26	1.24	.13	1.03	.2	1.03	.2	.92	.57	.64	5 B5
277	72	3.85	3.93	.37	.15	.81	-1.1	.75	-1.6	1.32	.75	.60	1 B1
283	72	3.93	4.02	.23	.15	.99	.0	1.05	.3	1.04	.68	.60	3 B3
288	72	4.00	4.09	.12	.15	.52	-3.3	.58	-2.8	1.51	.77	.59	4 B4
319	72	4.43	4.52	-.74	.19	.82	-.9	1.21	.9	.99	.40	.51	6 B6
331	72	4.60	4.67	-1.21	.21	1.64	2.5	1.96	3.1	.46	.07	.46	2 B2
288.3	72.0	4.00	4.06	.00	.16	.97	-.4	1.10	.1		.54		Mean (Count: 6)
31.9	.0	.44	.45	.79	.03	.34	1.8	.44	1.9		.25		S.D. (Population)
34.9	.0	.48	.50	.87	.03	.37	2.0	.48	2.1		.27		S.D. (Sample)

Model, Populn: RMSE .17 Adj (True) S.D. .77 Separation 4.67 Strata 6.56 Reliability .96  
 Model, Sample: RMSE .17 Adj (True) S.D. .85 Separation 5.14 Strata 7.18 Reliability .96  
 Model, Fixed (all same) chi-square: 131.5 d.f.: 5 significance (probability): .00  
 Model, Random (normal) chi-square: 4.8 d.f.: 4 significance (probability): .31

Analizden elde edilen güvenilirlik katsayısı 0,96'dır. Bu değer öğrencilerin sunum becerileri açısından sıralamasının oldukça yüksek bir güvenilirlikte elde edildiğini göstermektedir. Ayırma indeksi 4,67 ve güvenilirlik katsayısı 0,96 ile sabit etkiye ait “ öğrencilerin sunum becerileri arasında anlamlı farklılık vardır” hipotezi kay-kare ile test edildiğinde ( $\chi^2=131,5$ ,  $sd=5$ ,  $p=0,00$ ) reddedilmiştir. Bunun anlamı, öğrencilerin sunum becerileri düzeyleri birbirinden istatistiksel olarak anlamlı farklılık göstermektedir. Rasch analizi yüzeylere ait uygunluk içi ve uygunluk dışı değerlerini de üretmektedir. Bu değerlerin 0,5-1,5 arasında olması gerekmektedir (Wright, Linacre, Gustafson ve Martin-Lof, 1994). Uygunluk içi ve uygunluk dışı değerleri incelendiğinde sadece B2 kodlu bireye ait değerlerin belirlenen sınırların dışında kaldığı, bu öğrencinin belirlenen becerilerin dışında kaldığı ve uygun bir performans göstermediği söylenebilir.

### 3.2. Sunum Becerileri Değerlendirme Formu Analizi

Öğrencilerin sunum becerilerini değerlendirmek amacıyla kullanılan formda bulunan ölçütler ve bu ölçütlerin uygunluğu ile ilgili analiz sonuçları Tablo 2.'de verilmiştir.

**Tablo 2. Sunum Becerileri Değerlendirme Formu Analizi**

Total Score	Total Count	Obsvd Average	Fair (N) Average	Model Measure	Model S.E.	Infit InSq	Infit ZStd	Outfit InSq	Outfit ZStd	Estim. Discrim.	Correlation P/Res	PRExp	Nu Maddeler
127	36	3,53	3,63	,78	,20	,73	-1,2	,71	-1,3	1,31	,78	,70	3 Madde 3
136	36	3,78	3,90	,41	,21	,93	-,2	1,08	-,4	1,06	,70	,68	10 Madde 10
138	36	3,83	3,96	-,32	,21	1,11	,5	1,02	-,1	1,06	,77	,67	4 Madde 4
140	36	3,89	4,02	-,23	,21	,48	-,7	,84	-,5	1,15	,72	,67	2 Madde 2
140	36	3,89	4,02	-,33	,21	,98	-,0	1,04	-,2	,96	,66	,67	5 Madde 5
143	36	3,97	4,11	-,09	,22	,69	-1,3	1,05	-,2	1,14	,65	,66	9 Madde 9
144	36	4,00	4,14	-,04	,22	1,25	1,0	1,79	2,5	,37	,31	,66	1 Madde 1
147	36	4,08	4,22	-,11	,23	,85	-,5	,80	-,7	1,29	,78	,64	12 Madde 12
148	36	4,11	4,25	-,16	,23	,98	-,0	,88	-,3	1,07	,65	,64	7 Madde 7
153	36	4,23	4,39	-,44	,24	,76	-,9	,93	-,1	1,13	,67	,62	8 Madde 8
155	36	4,31	4,44	-,56	,25	1,14	-,3	1,00	-,1	,98	,59	,61	11 Madde 11
159	36	4,42	4,55	-,82	,26	1,39	1,3	2,02	2,5	,59	,35	,58	6 Madde 6
144,2	36,0	4,00	4,14	-,00	,23	,94	-,1	1,10	-,1		,64		Mean (Count: 12)
8,5	,0	,34	,24	-,43	,02	,23	1,1	,38	1,1		,15		S.D. (Population)
8,9	,0	,25	,25	-,45	,02	,26	1,1	,40	1,2		,15		S.D. (Sample)

odel, Populn: RMSE: .23 Adj (True) S.D.: .37 Separation: 1,62 Strata: 2,49 Reliability: .72  
odel, Sample: RMSE: .23 Adj (True) S.D.: .39 Separation: 1,72 Strata: 2,62 Reliability: .75  
odel, Fixed (all same) chi-square: 41,6 d.f.: 11 significance (probability): .00  
odel, Random (normal) chi-square: 8,7 d.f.: 10 significance (probability): .56

Analiz sonuçları incelendiğinde öğrencilerin sunum becerilerinde en zayıf kaldıkları noktaların sırayla Madde 3: Konuyu dinleyicilerin ilgisini çekecek şekilde sunma, Madde 10: Kaynak çeşitliliği ve Madde 4: Sunuyu hedefe yönelik materyalle destekleme olduğu görülmektedir. Öğrencilerin sunum becerilerinde en kolay sergiledikleri beceriler ise; Madde 6: Verilen sürede sunuyu yapma, Madde 11: Bilgisayar teknolojilerini kullanmadaki hâkimiyet olduğu görülmektedir.

Değerlendirme formunda bulunan ölçütlere ilişkin logit değerlerine ait standart hata (RMSE) değerinin 0,23 olduğu görülmektedir. Hesaplanan RMSE değerinin düşük olduğu, düzeltilmiş standart hata değeri (0,37) 1'in altında hesaplanarak desteklenmiştir. Ayırma indeksi 1,62 ve güvenilirlik katsayısı 0,72 ile sabit etkiye ait “ öğrencilerin sunum becerilerini belirlemede kullanılan ölçütlerin güçlükleri arasında anlamlı farklılık vardır” hipotezi kay-kare ile test edildiğinde ( $\chi^2=41,6$ ,  $sd=11$ ,  $p=0,00$ ) reddedilmiştir. Bir başka ifade ile sunum becerilerinin ölçülmesinde kullanılan ölçütlerin güçlükleri arasında anlamlı farklılık olduğu belirlenmiştir.

Yüzeylere ait uygunluk içi ve uygunluk dışı değerleri incelendiğinde ise bütün maddelerin uygunluk içi kriterini sağladığı, 6. ve 1. Maddelerin uygunluk dışı kriterini

sağlamadığı belirlenmiştir. Bu maddelerin sunum becerilerini ölçmede uygun maddeler olmadığı ifade edilebilir.

### 3.3.Puanlayıcı Analizi

Sunum becerilerini değerlendiren puanlayıcıların değerlendirme durumlarına ilişkin analiz sonuçları Tablo 3.'te verilmiştir. Puanlayıcıların katılık ve cömertlikleri incelendiğinde P2 kodlu puanlayıcının en cömert, P1 kodlu puanlayıcının en katı olduğu görülmektedir. Puanlayıcı katılık/cömertliklerine ilişkin RMSE değerinin 0,16 ve düzeltilmiş standart hata değerinin 0,76 olarak bulunması puanlayıcı analizine ilişkin standart hatanın düşük olduğunu gösterir.

**Tablo 3.** Puanlayıcı Analizi

Total Score	Total Count	Obsvd Average	Fair (M) Average	Model Measure	S.E.	Infit Mnsq	Zstd	Outfit Mnsq	Zstd	Estim. Discrn	Correlation PtMea	PtExp	N Puanlayıcı
321	72	4,46	4,54	2,29	,19	1,04	,2	1,41	1,7	,88	,42	,48	2 P2
318	72	4,42	4,50	2,19	,18	,81	-,9	,92	-,3	1,03	,46	,49	3 P3
315	72	4,38	4,46	2,09	,18	1,17	,8	1,62	2,6	,70	,27	,50	6 P6
290	72	4,03	4,12	1,41	,15	1,05	,3	1,01	,0	1,04	,62	,56	5 P5
244	72	3,39	3,45	,47	,14	,45	-4,4	,53	-3,5	1,61	,73	,62	4 P4
242	72	3,36	3,42	,43	,14	1,15	,9	1,09	,6	,93	,71	,62	1 P1
288,3	72,0	4,00	4,08	1,48	,16	,94	-,5	1,10	,2		,54		Mean (Count: 6)
33,6	,0	,47	,48	,78	,02	,25	1,9	,35	1,9		,16		S.D. (Population)
36,8	,0	,51	,52	,86	,02	,27	2,0	,38	2,1		,18		S.D. (Sample)

Model, Populn: RMSE ,16 Adj (True) S.D. ,76 Separation 4,65 Strata 6,54 Reliability ,96  
 Model, Sample: RMSE ,16 Adj (True) S.D. ,84 Separation 5,12 Strata 7,16 Reliability ,96  
 Model, Fixed (all same) chi-square: 149,3 d.f.: 5 significance (probability): ,00  
 Model, Random (normal) chi-square: 4,8 d.f.: 4 significance (probability): ,30

Ayrırma indeksi 4,65 ve güvenilirlik katsayısı 0,96 ile sabit etkiye ait “ puanlayıcıların katılık/cömertlikleri yönünden aralarında anlamlı farklılık vardır” hipotezi kay-kare ile test edildiğinde ( $\chi^2=149,3$ ,  $sd=5$ ,  $p=0,00$ ) reddedilmiştir. Bir başka ifade ile sunum becerilerini puanlayan puanlayıcıların katılıkları/cömertlikleri konusunda aralarında anlamlı farklılık olduğu belirlenmiştir. Puanlayıcılar arasında öğrenci performansları konusunda tutarsızlıklar bulunmaktadır. Yüzeylere ait uygunluk içi ve uygunluk dışı değerleri incelendiğinde ise bütün puanlayıcıların uygunluk içi kriterini sağladığı, 6. puanlayıcının uygunluk dışı kriterini sağlamadığı belirlenmiştir. .

## 4. SONUÇ VE TARTIŞMA

Bu çalışmanın amacı lisansüstü düzeyde yapılan bir ders kapsamında öğrenci sunum yapma becerilerine ait akran değerlendirme sonuçlarının çok yüzeyli Rasch ölçme modeli ile modelleyip sonuçları tartışmaktır.

Araştırmada öncelikle veri kalibrasyon haritası incelenmiştir. Buna göre sunum yapan bireylerin, becerilerin değerlendirilmesinde kullanılan kriterlerin ve puanlayıcıların aynı logit cetvel üzerinde sıralandıkları görülmüştür. Veri kalibrasyon haritası aynı ölçek üzerinde öğrenci, puanlayıcı ve maddeler arasındaki ilişkiye dair uygun bilgi edinmeyi

saplamaktadır (Nakamura, 2000). Araştırmada modellenen yüzeylerden biri biri olan puanlayıcı katılığı/cömertliğidir. Puanlayıcıların katılı/cömertlik karşılaştırmasına ilişkin RMSE (Root Mean Standart Error) değeri 0.23 bulunmuştur. Düzeltilmiş standart hata değerinin ise birin altında olduğu belirlenmiştir. Puanlayıcı ayırma indeksi de 1,62 olarak bulunmuştur. Bu değer, puanlayıcıların katılık/cömertlik düzeylerine göre farklılaştığını puanlayıcıların atadıkları puanlarda cömertlik/katılık hatasının olduğunu göstermektedir (Engelhardt ve Myford, 2003; Iramaneerat, Myford, Yudkowsky ve Lowenstein., 2009) İstenen değer puanlayıcılar arası uyum olması durumunda 0.00 ya da 0.00'a yakın değerlerdir (Myford ve Wolfe, 2004; Nakamura, 2000). Puanlayıcılar arası istenmeyen varyansın göstergesi olarak yorumlanan ayırma indeksi güvenilirliği (Engelhard, 2002) ise 0.72'dir. Bu değerler puanlayıcıların birbirleri yerine geçmelerinin sıkıntılı olduğu (Linacre, Wright ve Lunz, 1990) ve maddelerin niteliğinden ziyade puanlamayı yapan puanlayıcıya bağlı olduğunu (Sudweeks, Reeve ve Bradshaw, 2004) ifade etmektedir. Tüm bunlar doğrultusunda; bütün puanlayıcıların puanlama davranışlarının güvenilir olduğu, katılık ve cömertlik yönünden güvenilir bir şekilde sıralandıkları ve birbirlerinden farklılıklar gösterdiği görülmüştür. Çok yüzeyli rasch modelinin kullanıldığı çalışmalarda da (Atılğan, 2005a; Atılğan 2005b; Baştürk ve Işıkoğlu, 2007; Baştürk, 2008, Baştürk, 2009, Baştürk, 2010; Semerci, 2011a, Semerci, 2011b, Semerci 2011c, Akın ve Baştürk, 2012; Semerci, Semerci ve Duman, 2013; Yüzüak, Şahin ve Semerci, 2013; Yüzüak, Yüzüak ve Kaptan, 2015) puanlayıcıların bazen objektif bazen yanlı olabileceği belirtilmiştir.

Araştırmada kullanılan sunum becerileri değerlendirme formunda bulunan kriterlerin öğrencilerin sunum becerilerini ölçmek için amaca hizmet eder nitelikte olduğu belirlenmiştir. Çok yüzeyli Rasch ölçme modeli, diğer ölçme kuramları gibi güvenilirlik katsayısı üretmektedir. Elde edilen güvenilirlik katsayısının 1'e yaklaşması yüksek güvenilirliğe işaret etmektedir. Bu araştırmadan elde edilen güvenilirlik değerleri 0,72 ile 0,96 arasında değişmektedir. Bu değerler araştırmadan elde edilen sonuçların güvenilir olduğuna ve hataların azlığına işaret etmektedir. Bütün bu sonuçlar ışığında, çok yüzeyli Rasch ölçme modelinin bireyin performansının ölçülmesinde alternatif ölçme modeli olarak kullanılabilmesi bu örnek çalışma ile de desteklenmiştir.



## KAYNAKÇA

- Akın, Ö. ve Baştürk, R. (2012). Keman eğitiminde temel becerilerin Rasch ölçme modeli ile değerlendirilmesi. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 31(1), 175-187.
- Atılğan, H. (2005a). Genellenebilirlik kuramı ve puanlayıcılar arası güvenilirlik için örnek bir uygulama. *Eğitim Bilimleri ve Uygulama*, 4 (7), 95-108.
- Atılğan, H. (2005b). Müzik öğretmenliği özel yetenek seçme sınavının çok-yüzeyle rasch modeli ile analizi (İnönü üniversitesi örneği). *Eurasian Journal of Educational Measurement*, 20, 62 – 73.
- Baştürk, R. & Işıkoğlu, N. (2007). Okul öncesi eğitim kurumlarının işlevsel kalitelerinin çok-yüzeyle Rasch modeli ile analizi. *Kuram ve Uygulamada Eğitim Bilimleri*, 7(3), 727 – 752.
- Baştürk, R. (2008). Applying the many facet Rasch model to evaluate PowerPoint presentation performance in higher education. *Assessment and Evaluation in Higher Education*, 33(4), 431-444.
- Baştürk, R. (2009). Applying the many – facet Rasch model to evaluate powerpoint presentation performance in higher education. *Assesment and Evaluation In Higher Education*, 33(4), 431–444.
- Baştürk, R. (2010). Bilimsel araştırma ödevlerinin çok yüzeyle Rasch ölçme modeli ile değerlendirilmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(1), 51-57.
- Eckes, T. (2005). Examining rater effects in TestDAF writing and speaking performance assessments: A many facet Rasch analysis. *Language Assessments Quarterly*, 2, 197-221.
- Elhan, A. H. & Atakurt, Y. (2005). Ölçeklerin değerlendirilmesinde niçin rasch analizi kullanılmalı. *Ankara Üniversitesi Tıp Fakültesi Mecmuası*, 58, 47 - 50.
- Engelhard, G. (2002). Monitoring raters in performance assessments. In G. Tindal and T. Haladyna (Ed.), Large-scale assessment programs for all examinees: Validity, technical adequacy, and implementation (261–287). Mahwah, NJ: Lawrence Erlbaum.
- Engelhard, G. & Myford, C.M. (2003). *Monitoring faculty consultant performance in the Advanced Placement English Literature and Composition Program with a many-faceted Rasch model. ETS Research Report Series, (01)*. Princeton, NJ: Educational Testing Service.
- Iramaneerat, C., Myford, C.M., Yudkowsky, R. & Lowenstein, T. (2009). Evaluating the effectiveness of rating instruments for a communication skills assessment of medical residents. *Advances in Health Sciences Education*, 14 (4), 575-594.
- Linacre, J. M. (1993). Rasch-based generalizability theory. *Rasch Measurement Transactions*, 7(1), 283-284.
- Linacre, J.M. (1995). *Rasch measurement transaction*. MESA Press, Chichago, USA.
- Linacre, J. M. (1998). Structure in Rasch residuals: Why principal components analysis? *Rasch Measurement Transactions*, 12(2), 636.
- Linacre, J.M. (2002). Facets, factors, elements and levels. *Rasch Measurement Transactions*, 16, 880.
- Linacre, J.M., Wright B.D. & Lunz M.E. (1990). A Facets Model of Judgmental Scoring. Memo 61. MESA Psychometric Laboratory. University of Chicago. [www.rasch.org/memo61.html](http://www.rasch.org/memo61.html).

- Myford, C.M. & Wolfe, E.W. (2004). Detecting and measuring rater effects using many-facet Rasch measurement: Part II. *Journal of Applied Measurement*, 5, 189–227.
- Nakamura, Y. (2000). Many facet rasch based analsis of communicative language testing results. *Journal of Communication Students*, 12, 3-13.
- Popham, W.J. (2008). *Classroom assessment what teachers need to know*. USA: Pearson Education.
- Prieto, G. & Nieto, E. (2014). Analysis of rater severity on written expression exam using many faceted *Rasch measurement*. *Psicologica*, 35, 385-397.
- Semerci, Ç. (2011a, Elazığ). *Öğrencilerin BÖTE bölümüne ilişkin Rasch ölçme modeline gore değerlendirilmesi (Fırat Üniversitesi Örneği)*, 5th International Computer & Instructional Technologies Symposium'da sunulan sözlü bildiri, Fırat Üniversitesi, Elazığ.
- Semerci, Ç. (2011b). Mikro öğretim uygulamalarının çok-yüzeyle Rasch ölçme modeli ile analizi. *Eğitim ve Bilim/Education and Science*, 36(161), 14-25.
- Semerci, Ç. (2011c). Doktora yeterlikler çerçevesinde öğretim üyesi, akran ve öz değerlendirmelerin Rasch ölçme modeliyle analizi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 2(2), 164-171.
- Semerci, Ç., Semerci, N. & Duman, B. (2013). Yüksek lisans öğrencilerinin seminer sunu performanslarının çok-yüzeyle Rasch modeli ile analizi. *Sakarya Üniversitesi Eğitim Fakültesi Dergisi*, 25, 7-22.
- Sudweeks, R.R., Reeve, S. & Bradshaw, W.S. (2004). A comparison of generalizability theory and many-facet Rasch measurement in an analysis of college sophomore writing. *Assessing Writing*, 9 (3), 239-261.
- Wright, B.D., Linacre, M., Gustafson, J.E. & Martin-Lof, P. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370.
- Yüzüak, A.V., Şahin, A. E. & Semerci, Ç. (2013, Eylül). *Kimya laboratuvar uygulamalarının çok-yüzeyle Rasch ölçme modeliyle değerlendirilmesi*. 22. Ulusal Eğitim Bilimleri Kurultayı'nda sunulan sözlü bildiri, Eskişehir Osmangazi Üniversitesi, Eskişehir.
- Yüzüak, A. V., Yüzüak, B. & Kaptan, F. (2015). Performans görevinin akran gruplar ve öğretmen yaklaşımları doğrultusunda çok-yüzeyle rasch ölçme modeli ile analizi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6(1), 1-11.

## EXTENDED ABSTRACT

### 1. Introduction

Usage of the knowledge learned at school in daily life, developing the high level behaviors and assessing these behaviors come to the forefront in the last period of the education process. Paper-and-pencil tests are used for measuring the process data proceedings knowledge, skill and ability. But these tests are not enough for measuring skills and abilities. Therefore performance based measurement approach come into prominence. It enables to measure high level behaviors as researching, making a presentation and effective communication. In higher education one of the most frequently used skills is making presentation which aims to inform the others about a subject. Some of the utilities used to make presentation are Gold Disk, Astaund, Adobe Persuasion, Deltagraph Pro, Microsoft Powerpoint and Prezi.

One of the most important features for measuring a presentation is objectivity. Structured item format are not usually applied to students in the process. Therefore to use a grading scoring rubric is important. The graduator grade the individual's presentation according to the criteria in the scoring rubric, but decisions of graduator are not based only on the performance. The facets as difficulty of performance, characteristic of graduator and history of individual have also a place on scoring. The facets can be collected under some headings like influence of graduator, difficulty of criteria and criteria of scoring. This facet can influence the validity of scoring adversely.

Influence of graduator consisted of graduator characteristics that are not connected with performance. This source of variance cause error in measurement as it does not connected with the feature measured. This adverse effect can be brought under control by using some analysis method. Influence of graduator can be evaluated in the Many Facets Rasch Measurement Model.

Two theories in literature are used to explain latent features of the individual that take part. These are Classical Test Theory and Item Response Theory. Classical Test Theory has some limitations. Item Response Theory develops strong assumptions towards this limitations. Therefore researchers start to over utilize Item Response Theory. Item Response Theory is a measurement theory explains the latent variable underline item performance of individual by using the mathematical models.

Many Facet Rasch Measurement Model is the derivate of Rasch Model. Many Facet Rasch Measurement Model is different from Rasch model in respect to the number of facet and grade format. There is no limitation about number of facet in Many Facet Rasch Measurement Model. Besides, Many Facet Rasch Measurement Model enables polychotomous scoring.

In light of all of these factors, the aim of this study is to compose analyzed skills of making a presentation by using Many Facet Rasch Measurement Model. As part of this general aim,

- 1- Analyses of students' skills of making a presentation,
- 2- Analyses of generosity/parsimony of graduator,
- 3- Analyses of difficulty of task were made.

## 2. Method

This study was composed of six graduate students who took Measurement and Evaluation lesson at 2015-2016 academic year and one lecturer. “Skills of Making a Presentation Peer Assessment Form” developed by researcher in parallel with lesson outcomes were used for grading students’ skills of making presentation. Responses were categorized between “the most inadequate” and “the most adequate” as the five point Likert scoring. Facet software developed by Linacre (1993) was used for analyzing the data obtained in process of study. Three facets analyzed by using Rasch Model in this study are generosity/parsimony of gradulators, twelve criteria used in Skills of Making a Presentation Peer Assessment Form and seven gradulators.

## 3. Findings, Discussion and Results

According to the findings the student coded as B5 is the most successful one for presentation skills. The student coded as B2 is the most unsuccessful. The standard error value (RMSE) incidental to logit values of students’ skills of making a presentation is 0,17. The lowness of this value are supported with the values of adjusted standard error (0,77) calculated below 1. The reliability coefficient is 0,96. This value shows that the gradation of students’ skills was made reliably. The hypothesis as “there is statistically significant difference between students’ skills” was rejected by using Chi Square test ( $\chi^2=131,5$ ,  $sd=5$ ,  $p=0,00$ ). These findings mean unusually that there is statistically significant difference between students’ skills. According to proper and improper values, the student coded B2 didn’t perform well.

The most difficult items are item 3, item 10 and item 4 respectively. In addition to that the easiest items are item 6 and item 11. The standard error value (RMSE) incidental to logit values of difficulties of criteria is 0,23. The lowness of this value are supported with the values of adjusted standard error (0,37) calculated below 1. The reliability coefficient is 0,72. The hypothesis hypothesized as “there is statistically significant difference between difficulties of the criteria in peer form” was rejected by using Chi Square test ( $\chi^2=41,6$ ,  $sd=11$ ,  $p=0,00$ ). This findings’ mean is unusual that there is statistically significant difference between difficulties of the criteria in peer form. According to proper and improper values, item 1 and 6 were not proper for assessing students’ skills.

The most generous graduator was coded as P2. In addition to that the most parsimony graduator was coded as P1. The standard error value (RMSE) incidental to logit values of characteristic of gradulators is 0,16. The lowness of this value is supported with the values of adjusted standard error (0,76) calculated below 1. The reliability coefficient is 0,96. The hypothesis as “there is statistically significant difference between characteristics of gradulators” was rejected by using Chi Square test ( $\chi^2=149,3$ ,  $sd=5$ ,  $p=0,00$ ). This findings’ mean is unusual that there is statistically significant difference between characteristics of gradulators. According to the proper and improper values, the graduator coded P6 was not proper for assessing students’ skills.

The characteristics of gradulators were a source of variance for the scores of the individual. The gradulators participated in this study cannot take over the others. Besides, students’ skills of making a presentation can be assessed with the criteria in Skills of Making a Presentation Peer Assessment Form. The reliability coefficients are between 0,72 and 0,96. According to these coefficients the findings are reliable and the error level is low. All the findings show that Many Facet Rasch Measurement Model is a good alternative model for assessing skills and abilities.