

Genellenebilirlik Kuramında Dört Facetli Karışık Desen Kullanımı İçin

Örnek Bir Uygulama¹

Duygu ANIL*
Hacettepe Üniversitesi

Serap BÜYÜKKIDIK**
Hacettepe Üniversitesi

Özet

Bu araştırmada, performansa dayalı durum belirlemeden elde edilen ölçümlere Genellenebilirlik Kuramı'nda karışık desen uygulanmış ve dört facetli desende Genellenebilirlik ve Karar çalışmaları yürütülmüştür. Araştırma verileri 2011- 2012 bahar döneminde Kütahya ilinde bir ilköğretim okulunda 6-7-8. sınıf 132 öğrenciyle gerçekleştirilen performansa dayalı durum belirleme uygulamasından rastgele seçilen 90 öğrencinin sergiledikleri performansların dört puanlayıcı tarafından yanıt tanıma kodları ve göreve özgü analitik dereceli puanlama anahtarını kullanarak puanlanmasından elde edilmiştir. Araştırma sonucunda, ölçmenin nesnesi olan bireyin en fazla toplam varyansı açıklama yüzdesine sahip olduğu, sınıf düzeyinin önemsenmeyecek düzeyde değişkenlik kaynağı olduğu, bunun yanında farklı sınıflardaki bireylerin puanlarının puanlayıcıdan puanlayıcıya kısmen değişebileceği görülmüştür. Yapılan Karar çalışması sonucunda ise G katsayısının 0.911, phi katsayısının ise 0.898 olduğu ve puanlayıcı sayısını artırıp azaltmanın; görev sayısını artırıp azaltmaya oranla daha fazla ölçümün güvenilirliğine etkisi olduğu bulunmuştur.

Anahtar kelimeler: Genellenebilirlik Kuramı, karışık desen, dereceli puanlama anahtarı

Abstract

In this research has been used mixed design in Generalizability Theory in performance assessment applications' scores and has been made Generalizability and Decision Study for analyzing reliability of the scores. Research data has collected the 132 second tier students randomly selected 90 students at a primary school in Kütahya in 2011-2012 spring term. According to the results of the study, person which is subject of measurement has the highest the percentage of total variance description, has seen insignificant as a source of variability in the level of class, as well as the scores of different classes can change partly seen from scorer to scorer. As a result of the Decision Study has been 0.911 G coefficient, phi coefficient that the 0.898, and increase or decrease the number of raters, increasing the number of tasks to reduce the effect was more than the reliability of the measurement.

Key words: Generalizability Theory, mixed design, rubric

Alanyazındaki Genellenebilirlik Kuramı uygulamaları incelendiğinde; sağlık bilimlerinden (Kozaki, 2004; Jarjoura et al, 2004; Al-Mahroos, 2009), müzik eğitimine (Atılğan, 2004; Bergee;

¹ Bu araştırma, Abant İzzet Baysal Üniversitesi'nde düzenlenen III. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi'nde (19-21 Eylül 2012) sunulmuştur.

* Doç. Dr., Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü Eğitimde Ölçme ve Değerlendirme Bilim Dalı, aduygu@hacettepe.edu.tr

** Yüksek Lisans Öğrencisi, Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü Eğitimde Ölçme ve Değerlendirme Bilim Dalı, sbuyukkidik@gmail.com

2007), davranış bilimlerinden (Hoyt ve Melby, 1999; Tindal et. al, 2008; Christ et. al., 2010), insan kaynaklarına (Van Hooft et. al., 2006; Arce-Ferrer ve Castillo, 2007; Yelboğa, 2007) kadar elde edilen ölçümlerin güvenilirliğinin sınanmasında bir çok alanda karşımıza çıkmaktadır. Yaygın kullanım alanına sahip olan Genellenebilirlik Kuramı, bir ölçme durumunda yer alan bütün potansiyel hata kaynaklarını birlikte ve eş zamanlı olarak değerlendirerek, tek bir hata kaynağını bir defada değerlendiren Klasik Test Kuramı'na göre daha kapsamlı bir güvenilirlik kestirimi yapılmasına olanak sağlaması (Shavelson ve Webb, 1991; Brennan, 2001) yönüyle ülkemizde yapılan araştırmalarda son zamanlarda sıklıkla kullanılan kuramlardan biri olmuştur. Genellenebilirlik Kuramında ölçümlerin özelliklerine göre farklı desenler bulunmaktadır.

Genellenebilirlik Kuramında Kullanılan Desenler ve Facet Kavramı

Genellenebilirlik Kuramında değişkenlik kaynağının sayısına bağlı olarak desenin oluşturulmasının yanı sıra, çaprazlanmış (crossed) ya da yuvalanmış (nested) olmak üzere iki türden desen vardır. Ölçmedeki değişkenlik kaynaklarının bütün koşulları diğer bir değişkenlik kaynağının bütün koşullarını etkiliyorsa çaprazlanmış ve değişkenlik kaynakları arasında "x" işareti konularak gösterilir. Bir değişkenlik kaynağının bazı koşulları, diğer bir değişkenlik kaynağının bazı koşullarınca gözlemleniyorsa yuvalanmıştır ve değişkenlik kaynakları arasında ":" konularak gösterilir (Shavelson ve Webb, 1991; Brennan, 2001; Mushquash ve O'Connor, 2006). Bazı değişkenlik kaynaklarının çaprazlanması, bazılarının yuvalanması durumunda ise karışık desen mevcuttur. Facet ölçme hatasının olası kaynağı olarak da tanımlanabilir (Brennan, 2001). Bu çalışmada facet kavramının Türkçe karşılığı gibi kullanılan değişkenlik kaynağı kavramının tercih edilmeme nedeni araştırmadaki desende ölçmenin nesnesi kabul edilen bireyin de bir değişkenlik kaynağı olup; ancak ölçme hatasının olası kaynağı teşkil etmemesi sebebiyle facet sayılmamasıdır.

Ülkemizde Genellenebilirlik Kuramı İle Yapılan Çalışmalar

Ülkemizde yapılan Genellenebilirlik Kuramı ile ilgili çalışmalar incelendiğinde genelde birey/öğrenci, görev/madde ve puanlayıcı/hakem/gözlemci değişkenlik kaynaklarının ele alındığı ve en fazla iki facetli evrenler üzerinde çalışma yapıldığı görülmektedir (Atılğan, 2004; Yelboğa, 2007; Güler, 2008; Deliceoğlu, 2009; Nalbantoğlu, 2009). Sosyal bilimlerdeki araştırmalarda en önemli geçerlik sorunlarından biri ölçmenin amacı dışındaki dışındaki değişkenlik kaynaklarının yapılan ölçümlere etki etmesidir. Temelinde Anova'ya dayanan Genellenebilirlik Kuramı ile araştırmamızda tanımladığımız tüm değişkenlik kaynaklarının birlikte ve tek başına toplam varyansı açıklama yüzdesi elde edilebilmektedir

Bu amaçla bu çalışmada; birey, sınıf, puanlayıcı, görev, ölçüt değişkenlik kaynaklarını ele alan dört facetli karışık desen uygulamasında Genellenebilirlik ve Karar çalışması bulguları incelenecektir.

Yöntem

Araştırmanın Türü

Araştırma, psikometrik özelliklerin belirlenmesine ve durum saptamaya yönelik olup, genelleme amacı gütmemesi boyutuyla betimseldir. Genellebilirlik Kuramında karışık desen uygulamasını içermesi yönüyle de temel araştırmadır.

Verilerin Toplanması

Araştırmanın verileri, 2011- 2012 eğitim-öğretim yılında Kütahya İline bağlı merkez ilköğretim okulunda öğrenim gören 132 ilköğretim ikinci kademe 6.,7.,8. sınıf öğrencisinin rutin olmayan problem çözme becerisine yönelik hazırlanan iki performans görevinde sergiledikleri performansların dört puanlayıcı tarafından puanlanması ile toplanmıştır. Kullanılan desen açısından her bir sınıftaki öğrencilerin sayısının eşit olması için 6. ve 7. sınıflardan bilgisayarda rastgele 30 öğrenci seçilip, her sınıftan 30 öğrenci olacak şekilde 90 öğrencinin performans puanları üzerinden veriler analiz edilmiştir.

Verilerin Analizi

Ölçmenin nesnesi olarak kabul edilen farklı sınıf düzeylerindeki (s) 90 bireyin (b), iki görevde (g) gösterdikleri performansların analitik dereceli puanlama anahtarında yer alan üç ölçüt (ö) çerçevesinde dört puanlayıcı (p) tarafından puanlanmasından elde edilen puanlara “(b : s) x p x ö x g” karışık deseni uygulanmıştır. Araştırmanın birinci aşamasında, Genellebilirlik Kuramında G çalışması yürütülerek ana ve ortak etkiler için “(b : s) x p x ö x g” karışık deseninde varyans değerleri kestirilmiştir. İkinci aşamada ise aynı desen için Karar çalışması yürütülerek puanlayıcı ve görev sayısının bir arttırılıp azaltılması durumunda G ve phi katsayısı kestirimlerine yer verilmiştir. Verilerin analizinde SPSS 15.0 ve EduG 6.0 e programından yararlanılmıştır.

Bulgular

Bu araştırmada, ilköğretim ikinci kademedeki her üç sınıftan (s) seçilen ölçmenin nesnesi kabul edilen 90 bireyin (b), iki görevde (g) gösterdikleri performanslar, dört puanlayıcı (p) tarafından, analitik dereceli puanlama anahtarındaki üç ölçüt (ö) çerçevesinde puanlatılmasından elde edilen verilere (b : s) x p x ö x g dört facetli karışık desen kullanılmıştır. Tablo 1’de yapılan Genellebilirlik çalışması sonucu değişkenlik kaynaklarının varyans bileşenleri bulunmaktadır.

Tablo 1. (b:s) x p x ö x g Deseni İçin G Çalışması ile Ölçmenin Kestirilen Varyansları ve Toplam Varyansı Açıklama Oranları

Değişkenlik kaynağı	Toplam kareler	Sd	Bileşenler			
			Kareler ortalaması	Varyans	Hata	%
s	36.173	2	18.086	-0.046	0.021	0.0
b:s	3578.013	87	41.126	1.582	0.257	64.0
p	125.187	3	41.729	0.054	0.050	2.2
ö	16.806	2	8.403	0.011	0.008	0.4

g	18.890	1	18.890	0.016	0.014	0.7
sp	72.104	6	12.017	0.051	0.033	2.1
sö	0.632	4	0.158	0.000	0.001	0.0
sg	2.512	2	1.256	0.004	0.002	0.2
bp:s	763.875	261	2.926	0.446	0.042	18.1
bö:s	35.394	174	0.203	0.011	0.004	0.5
bg:s	30.930	87	0.355	0.021	0.005	0.9
pö	6.793	6	1.132	0.001	0.004	0.1
pg	0.564	3	0.188	-0.000	0.002	0.0
ög	0.084	2	0.042	-0.000	0.001	0.0
spö	6.989	12	0.582	-0.003	0.006	0.0
spg	1.210	6	0.201	-0.006	0.003	0.0
sög	0.287	4	0.071	-0.004	0.002	0.0
bpö:s	132.050	522	0.252	0.002	0.010	0.1
bpg:s	62.891	261	0.240	-0.002	0.008	0.0
bög:s	18.294	174	0.105	-0.035	0.004	0.0
pög	6.137	6	1.022	0.002	0.006	0.1
spög	9.378	12	0.781	0.017	0.009	0.7
bpög:s,e	129.816	522	0.248	0.248	0.015	10.0
Toplam	5055.02037	2159				100%

Tablo 1 incelendiğinde, Genellenebilirlik Kuramında “(b : s) x p x ö x g” deseninde ölçmenin nesnesi olan farklı sınıflardaki bireylerin değişkenlik kaynağının varyansı $\sigma_{b:s}^2$ (1.582), toplam varyansı açıklama yüzdesi % 64 ile en yüksek değerlerde bulunmuştur. Farklı sınıflardaki bireylerin farklı puanlayıcılardan almış olduğu puanların değişkenlik kaynağının varyansı $\sigma_{(b:s)p}^2$ (0.446), toplam varyansı açıklama yüzdesi % 18.1; artık olarak da adlandırılan tesadüfi hata olarak kabul edilen tüm değişkenlik kaynaklarının ortak etkisinin varyansı $\sigma_{(b:s)pog}^2$ (0.248), toplam varyansı açıklama yüzdesi % 10.0; puanlayıcı değişkenlik kaynağının varyansı σ_p^2 (0.054), toplam varyansı açıklama yüzdesi % 2.1 bulunmuştur. Bunun dışında sınıf düzeyi σ_s^2 (0.000) varyans miktarı ve % 0.0 toplam varyansı açıklama yüzdesi ile önemli bir değişkenlik kaynağı olmamıştır.

Tablo 2. Performansa Dayalı Durum Belirlemeden Elde Edilen Puanların (b:s) x p x ö x g Deseninde Puanlayıcı ve Görev Sayılarının Artırılıp Azaltılmasıyla Yapılan Karar Çalışması (K) Sonuçları

ng	np	G	Phi	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
1	4	0.899	0.882	0.176	0.211
2	3	0.888	0.873	0.198	0.228
2	4	0.911	0.898	0.153	0.178
2	5	0.926	0.914	0.125	0.148
3	4	0.916	0.904	0.145	0.168

ng : görev sayısı, np : puanlayıcı sayısı

Not: Koyu yazılmış olanlar çalışmada kullanılan puanlayıcı ve görev sayılarıdır.

Performansa dayalı durum belirlemeden elde edilen puanların (b:s) x p x ö x g deseninde G katsayısı 0,911; Phi katsayısı ise 0.898 olarak kestirilmiştir. Yapılan karar çalışması sonucunda, Tablo 2. incelendiğinde görev ve puanlayıcı sayısını arttırmanın G ve Phi katsayısını arttırdığı, azaltmanın ise G ve Phi katsayısını azalttığı görülmektedir.

Ayrıca puanlayıcı sayısı sabit tutularak görev sayısının bir arttırılmasında ($n_p = 4, n_g = 3$) G katsayısı 0.916 olup, 0.005 artarken; phi katsayısı 0.904 olup, 0.006 artmaktadır. Görev sayısı sabit tutulup, puanlayıcı sayısı bir arttırıldığında ($n_p = 5, n_g = 2$) ise; G katsayısı 0.926 olup, 0.015 artarken; phi katsayısı 0.914 olup, 0.016 artmaktadır. Puanlayıcı sayısı sabit tutularak görev sayısının bir azaltılmasında ($n_p = 4, n_g = 1$) G katsayısı 0.899 olup, 0.012 azalırken; phi katsayısı 0.882 olup, 0.016 azalmaktadır. Görev sayısı sabit tutulup, puanlayıcı sayısı bir azaltıldığında ($n_p = 3, n_g = 2$) ise; G katsayısı 0.888 olup, 0.023 azalırken; phi katsayısı 0.873 olup, 0.025 azalmaktadır. Bu sonuçlardan ölçümlerin güvenilirliğini etkilemede puanlayıcı sayısını değiştirmenin, görev sayısını değiştirmeye oranla etkili olduğu anlaşılmaktadır. Aynı şekilde puanlayıcı ya da görev sayısını bir azaltmanın, bir arttırmadan daha etkili olduğu da Tablo 2. incelendiğinde görülmektedir.

Sonuç ve Öneriler

Yapılan Genellenebilirlik çalışması sonucunda en büyük değişkenlik kaynağı ölçmenin nesnesi birey çıkmıştır. Bu sonuç uygulama yapılan grubun ölçülen özellik bakımından heterojen olduğunun göstergesidir. Bireylerin durumlarının bir puanlayıcıdan diğerine kısmen farklılık gösterdiği, en büyük ikinci (b:s) x p varyans oranıyla anlaşılmaktadır. Artık etki olarak kabul edilen bütün değişkenlik kaynaklarının ortak etkileşimi ise en büyük üçüncü değişkenlik kaynağı olmuştur. Ayrıca yapılan ölçümde görev, sınıf ve alt ölçütün önemsenmeyecek büyüklükte toplam varyansa katkısı olduğu da görülmektedir. Aynı görevlerin farklı sınıflara uygulanması sonucunda sınıf düzeyi önemli bir değişkenlik kaynağı olmamıştır; bunda üst sınıfların performans görevlerini yerine getirirken güdülenme düzeyinin alt sınıflara göre daha düşük olmasının ya da uygulama yapılan grupta alt ve üst sınıflarda benzer oranlarda üst düzey düşünme becerilerine sahip bireylerin olmasının etkisi olabileceği düşünülmektedir. Puanlayıcı değişkenlik kaynağı ise küçük bir oranda toplam varyansı etkilemektedir, bu puanlayıcıların kısmen birbirleriyle tutarlı puanlama yaptıklarının göstergesi olabilir.

Karar çalışması sonucunda ise en yüksek G ve phi katsayısı değerlerine görev sayısı sabit tutulup, puanlayıcı sayısı bir arttırıldığında ($n_p = 5, n_g = 2$) ulaşıldığı görülmüştür. Her durumda, G katsayısı phi katsayısından göreceli olarak büyük çıkmıştır. Puanlayıcı sayısını arttırmanın, görev sayısını arttırmadan daha fazla ölçümlerin güvenilirliğine katkı getirdiği de görülmüştür.

İleride yapılacak araştırmalarda farklı değişkenlik kaynaklarının alındığı, karışık desenler kullanılabilir. Ayrıca birden fazla desen çalışılıp, değişkenlik kaynaklarının ölçümlerin güvenilirliğine etkisi karşılaştırılabilir.

Puanlayıcı kanaatlerinin devreye girdiği durumlarda mümkünse göreve özgü ölçme araçları hazırlanmalı, davranış tanıma kodu çıkartılarak yapılan ölçümler nesnellığe yakınlaştırılmaya çalışılmalıdır.

Kaynaklar

- Al-Mahroos, F. (2009). Construct Validity and Generalizability of Pediatrics Clerkship Evaluation at a Problem-Based Medical School, Bahrain. *Evaluation & the Health Professions*, 32 (2), 165-183.
- Arce-Ferrer, A. J., Castillo, I. B. (2007). Investigating Postgraduate College Admission Interviews: Generalizability Theory Reliability and Incremental Predictive Validity. *Journal of Hispanic Higher Education*, 6: 118.
- Atılğan, H. (2004). *Genellenebilirlik Kuramı ve çok değişkenlik kaynaklı Rasch modelinin karşılaştırılmasına ilişkin bir araştırma*. Yayınlanmamış doktora tezi. Hacettepe Üniversitesi: Ankara.
- Brennan, R. L. (2001). *Generalizability Theory*. Springer-Verlag New York, Inc.
- Christ, T. J., Tillman C., Chafouleas, S. M., & Boice C. H. (2010). Direct Behavior Rating (DBR): Generalizability and dependability across raters and observations. *Educational and Psychological Measurement*. 70(5), 825–843.
- Deliceoğlu, G. (2009). *Futbol yetilerine ilişkin dereceleme ölçeğinin genellenebilirlik ve klasik test kuramına dayalı güvenilirliklerinin karşılaştırılması*. Yayınlanmamış doktora tezi. Ankara: Ankara Üniversitesi.
- Güler, N. (2008). *Klasik Test Kuramı Genellenebilirlik Kuramı ve Rasch Modeli üzerine bir araştırma*.Yayınlanmamış doktora tezi. Hacettepe Üniversitesi: Ankara.
- Hoyt, W. T. ve Melby, J. N. (1999). Dependability of measurement in counseling psychology : an introduction to Generalizability Theory. *The Counseling Psychologist*, 27: 325.
- Jarjoura, D., Early, L., & Androulakakis, V. (2004). A multivariate generalizability model for clinical skills assessments. *Educational and Psychological Measurement*, 64: 22.
- Kozaki, Y. (2004). Using GENOVA and FACETS to set multiple standards on from japanese into english performance assessment for certification in medical translation. *Language Testing*, 21: 1.
- Martin J. Bergee. (2007). Performer, rater, occasion, and sequence as sources of variability in music performance assessment. *Journal of Research in Music Education*, 55: 344-358
- Mushquash, C. ve O'Connor, B. P. (2006). SPSS and SAS programs for Generalizability Theory analysis. *Behavior Research Methods*. 38 (3), 542-547.
- Nalbantoğlu, F. (2009). *Performans ölçümlerinde Genellenebilirlik Kuramıyla farklı desenlerin karşılaştırılması*.Yüksek lisans tezi: Hacettepe Üniversitesi: Ankara.
- Shavelson, R. J. ve Webb, N. M. (1991). *Generalizability Theory: a primer*. Sage Publications, USA.
- Tindal, G., Yovanoff, P., & Geller, J.P. (2008). Generalizability theory applied to reading assessments for students with significant cognitive disabilities. *J Spec Educ*, 44: 3.
- Van Hooft, E. J. A.; Born, M.; Taris, T.W., & Van Der Flier, H. (2006). The cross-cultural generalizability of the theory of planned behavior : a study on job seeking in the Netherlands. *Journal of Cross-Cultural Psychology*, 37: 127.
- Yelboğa, A. (2007). *Klasik Test Kuramı ve Genellenebilirlik Kuramına göre güvenirlüğün bir iş performansı ölçeği üzerinde incelenmesi*. Doktora Tezi. Ankara Üniversitesi: Ankara.