

Makalenin Türü / Article Type : Araştırma Makalesi / Research Article
Geliş Tarihi / Date Received : 18.09.2019
Kabul Tarihi / Date Accepted : 10.03.2020
Yayın Tarihi / Date Published : 02.06.2020



[doi https://dx.doi.org/10.17240/aibuefd.2020..-621529](https://dx.doi.org/10.17240/aibuefd.2020..-621529)

ÖRTÜK SINIF ANALİZİNİN FARKLI PUANLAMA DURUMLARINDA İNCELENMESİ

Hüseyin SELVİ¹, Didem DERİCİ YILDIRIM²

ÖZ

Bu çalışmada Örtük Sınıf Analizinin farklı puanlama durumlarındaki işleyişini incelemek amaçlanmıştır. Bu amaç doğrultusunda çoktan seçmeli maddelerden oluşan ve seçenekleri ikili, uzman yargısına dayalı ağırlıklı ve deneysel ağırlıklı puanlanabilen bir test kullanılarak Mersin Üniversitesi Eğitim Fakültesi'nde öğrenim görmekte olan toplam 595 öğrenciden veri toplanmıştır. Öğrencilerin test maddelerine vermiş oldukları yanıtlar ikili, deneysel ağırlıklı ve uzman yargısına dayalı ağırlıklı olarak ayrı ayrı puanlanmış ve elde edilen veriler üzerinde örtük sınıf analizi yapılmıştır. Ulaşılan bulgular ikili ve uzman yargısına dayalı ağırlıklı puanlama için örtük sınıf analizinde aynı sınıf sayısına ulaşıldığını göstermiştir. En az sınıf sayısına ise en az parametre kestirimiyle deneysel ağırlıklı puanlama için ulaşılmıştır. İkili ve uzman yargısına dayalı ağırlıklı puanlama için ise elde edilen sınıf sayısı ve kestirilen parametre sayısı deneysel ağırlıklı puanlamayla ulaşılandan daha yüksektir. Örtük sınıf analizinde en uygun model, en az örtük sınıf sayısı ve en az parametre kestirimiyle model veri uyumunu yakalayan modeldir. Bu nedenle deneysel ağırlıklı puanlama yöntemiyle ulaşılan modelin örtük sınıf analizi için en uygun model olduğu ifade edilebilir.

Anahtar Kelimeler: Örtük sınıf analizi, ikili puanlama, deneysel ağırlıklı puanlama, uzman yargısına dayalı ağırlıklandırma

INVESTIGATION OF LATENT CLASS ANALYSIS IN DIFFERENT SCORING SITUATIONS

ABSTRACT

The aim of this study is to examine the Latent Class Analysis in different scoring situations. For this purpose, a test consists of multiple choice items which can be scored in different types such as dichotomous, expert judgement and weighted experimentally, respectively. A total of 595 students who are studying at Mersin University in the Faculty of Education were included in the study. Latent Class Analysis was used to determine the best fitted model of data. The findings of the study showed that the same number of classes was calculated in latent class analysis for dichotomous and expert judgment scoring. The minimum number of classes was obtained for experimentally scoring with the least parameter estimation. For dichotomous and expert judgment scoring, the number of classes was higher than experimental scoring. In the latent class analysis, the most appropriate model is the best fitted model with the least number of latent classes and least parameter estimation. Therefore, it can be stated that the experimental scoring method model is the most appropriate model for latent class analysis.

Keywords: Latent class analysis, dichotomous scoring, expert judgment scoring, experimentally weighted scoring

¹ Mersin Üniversitesi, Tıp Fakültesi, Tıp Eğitimi AD, hsyn_selvi@yahoo.com.tr, <https://orcid.org/0000-0002-3513-0003>

² Mersin Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim AD, didemderici@hotmail.com, <https://orcid.org/0000-0001-7709-6133>

1.GİRİŞ

Sosyal bilim arařtırmalarında gruplar arası karřılařtırmalara sıklıkla başvurulmaktadır. Grupların ve grup üyeliđinin bilindiđi durumlarda bu amaçlar kolaylıkla gerçekteřirilebilmektedir. Ancak karřılařtırma yapılacak gruplara iliřkin grup bilgisi her zaman elde edilmemiř ya da dođrudan gözlenemiyor (örtük) olabilir. Bu durumda popülasyon içinde yer alan bazı bireylerin diđerlerinden farklı davranıřlar sergilemesi elde edilmemiř, öngörülemediř ya da dođrudan gözlenemeyen bir grubun ya da grupların varlıđından kaynaklanıyor olabilir (MacDonald, 2018).Böyle bir durumda gerek bireylerin gerekse deđiřkenlerin örtük bir gruba ya da yapıya ait olup olmadıđının ortaya konması çalıřmalarda güvenilir ve geçerli sonuçlar ortaya konabilmesi açařından önemlidir.

Örtük grupların incelenmesinde ise sıklıkla örtük deđiřken modellerinden (latent variable models) yararlanılmaktadır (Güngör Çulha ve Korkmaz, 2011). Örtük deđiřken modelleri ‘faktör analizi’, ‘madde tepki kuramı’, ‘örtük profil analizi’, ve ‘örtük sınıf analizi’ olmak üzere 4 bařlık altında sınıflandırılmıřtır (Bartholomew, Knott ve Moustaki, 2011). Bu dört bařlık arasındaki en önemli ayırım ise deđiřkenlerden elde edilen verilerin ölçme düzeylerine dayanmaktadır. Faktör analizi ve örtük özellik analizinde örtük deđiřkenlerden sürekli yapıda veri elde edilebilirken, örtük profil ve örtük sınıf analizinde ise kesikli ya da sıralamalı düzeyde veri elde edilebilmektedir.

Bunlardan ‘örtük sınıf analizi’ (ÖSA), ikili (sürekli) yapıdaki gözlenen deđiřkenlerden yararlanarak bireyleri sınıflamak amacıyla Lazarsfeld (1968) tarafından geliřtirilmiřtir. ÖSA’da bireylerden elde edilen veriler kullanılarak bireylerin örtük sınıflara atanma olasılıkları istatistiksel olarak hesaplanmaktadır. ÖSA ortaya atıldıktan sonra, Goodman (1974) tarafından ikiden fazla kategorisi olan deđiřkenler için model parametrelerinin maksimum olabilirlik tahminleriyle elde edilebildiđi bir algoritma daha geliřtirilmiřtir. Sonraki yıllarda Clogg (1988) kovaryantlar, gözlenen deđiřkenler arasındaki iliřki ve deđiřken tipine göre ÖSA’yı daha da geliřtirmiřtir. Hagenars (1990) ve Vermunt (1997) ise, ÖSA’nın günümüzdeki çerçevesini oluřturmuřtur.

Örtük sınıf analizi en basit haliyle, popülasyonun çok sayıda gözlenemeyen gruptan oluřtuđunu varsaymakta olup, bireysel farklılıkları formülize etmektedir. ÖSA’da Yij bir bireyin i. maddeden j.maddeye (i=1,...I ve j=1.....J) kadar vermiř olduđu yanıtlar olmak üzere, verilen yanıtlardan yararlanılarak tahmin edilen örtük sınıf sayısı (C) hesaplanmaktadır. Bu hesaplamayı yapabilmek için kiřinin cevaplarının dođru ya da yanlıř olmasının herhangi bir önemi yoktur. Elde edilen her bir örtük sınıf zi=1’den C’ye olmak üzere zi ile tanımlanmaktadır. Her bir örtük sınıfın homojenliđi ise kořullu bağımsızlık varsayımı ile sađlanmaktadır. Bu varsayım altında, marjinal olasılıklar Eřitlik 1’den yararlanarak hesaplanmaktadır.

$$P(y_i) = \sum_{c=1}^C P(y_i | z_i = c) P(z_i = c) \quad (1)$$

Elde edilen marjinal olasılıklardan yola çıkarak Eřitlik 2’de ki formülden ise sonsal olasılıklar elde edilmektedir.

$$P(z_i | y_i) = \frac{P(y_i | z_i) P(z_i)}{P(y_i)} \quad (2)$$

Sonsal olasılık deđerleri 0 ile 1 arasında deđiřmektedir ve bir bireyin diđer sınıflarla karřılařtırıldıđında ait olduđu sınıfı olasılıksal olarak göstermektedir (Rijmen, 2011).

Örtük sınıf analizinde bireylerin sınıflandırılmasının yanı sıra, kullanılan veriye en uygun model bulunarak en uygun sınıf sayısının belirlenmesi amaçlanmaktadır. Bunun için kullanılan bazı kriterler mevcuttur. Bu kriterlerden en sık kullanılanları Akaike Bilgi Kriteri (AIC) ve Bayes Bilgi Kriteri (BIC)’dir. Her ikisi için de en küçük deđerin elde edildiđi modelin en iyi olduđuna karar verilir. Bu modele ait sınıf sayısı da en uygun sayı olarak belirlenmektedir. ÖSA iteratif bir yöntem olup, sınıf sayısının 1 olduđu durumdan bařlanmakta ve her adımda sınıf sayısı arttırılmaktadır (Madigson ve Vermunt, 2004).

ÖSA’da ve diđer örtük deđiřken modellerinde gözlenen deđiřkenlerden hareketle örtük deđiřkenler güvenilir ve geçerli řekilde ortaya konmaya çalıřılmaktadır. Ancak ÖSA’da dahil olmak üzere kullanılan örtük deđiřken modellerinin hemen tamamının güçlü ve zayıf yanlarının olduđu bilinmektedir. Alanyazında bu modellerin güçlü ve zayıf yanlarını ortaya koymaya yönelik pek çok çalıřma yapılmıřtır. Yapılan bu çalıřmalar farklı örtük deđiřken modelleriyle elde edilen bulguların, örnekleme büyüklüđü, madde sayısı, test uzunluđu, gibi deđiřkenlerden etkilenebildiđini ortaya koymaktadır (Galindo-Garre ve Vermunt, 2006; Tueler, Drotar ve Lubke, 2011; Wurpts ve Geiser, 2014).

Diđer taraftan sosyal bilimler için her maddenin ve seçeneđin psikolojik uzaydaki karřılıklarının eřit olmasını beklemek mümkün deđildir. Bu nedenle ÖSA’dan elde edilen bulguları deđiřtirebilecek bir başka deđiřkenin ise kullanılan ölçme aracında yer alan maddelerin puanlanma řekli olduđu düşünölmektedir. Eđitim ve psikoloji bilimleri bařta olmak üzere bireylerin örtük özelliklerine (latent traits) iliřkin dođrudan gözlenemeyen

değişkenlerin sıklıkla çalışıldığı bilim dallarında, değişkenlerin fiziksel büyüklüklerini anlayabilmek için bireylerin gözlenebilen çeşitli davranışlarından ve tepkilerinden hareketle çeşitli çıkarsamalar yapılmaktadır (Anastasi ve Urbina, 1997). Değişkenlerin doğrudan gözlenebildiği fiziksel büyüklüklerin ölçülmesinde bu çıkarsamalar kolay bir şekilde gerçekleştirilebilirken, doğrudan gözlenemeyen değişkenlerin ölçülmesinde ise oldukça karmaşıktır. Burada gözlenen davranış ve tepki örüntülerinin ne kadarlık miktarına ne kadar puan atfedileceği konusu ise bireylerin örtük özelliklerinin ortaya konmasında son derece önemlidir. Bu amaçla alan yazında geliştirilen farklı puanlama yöntemleri bulunmaktadır. Sıklıkla kullanılan puanlama yöntemlerinden biri de ikili (dichotomous) puanlamadır. İkili puanlama (1-0), sağlamış olduğu puanlama kolaylığı nedeniyle başta çoktan seçmeli test maddelerinin puanlanmasında sıklıkla tercih edilmektedir. İkili puanlamada, genellikle doğru yanıtlara '1' puan, doğru yanıt dışındaki durumlara ise '0' puan verilmektedir (Gulliksen, 1950). Diğer bir ifadeyle maddeyi bilerek ya da şans başarısıyla doğru cevaplayanların 'tam bilme'; maddeyi boş bırakan, yanlış cevap veren, birden fazla seçenek işaretleyen, süre kısıtlaması nedeniyle maddeye erişemeyen, doğru yanıtı kısmen bilen, cevap formunda kaydırma yapanların, vb. ise 'tam yanlış bilme' durumuna sahip olduğu varsayılmaktadır (Embretson ve Reise, 2000). Bu varsayım nedeniyle ikili puanlama, alanyazında bireylerin gerçek bilgi düzeylerini yansıtmakta yetersiz kaldığı (Embretson ve Reise, 2000) gerekçesiyle eleştirilmektedir. Bu eleştirilere çözüm üretmek adına maddelerin ya da seçeneklerin farklı şekillerde ağırlıklandırıldığı, farklı ağırlıklı puanlama yöntemleri geliştirilmiştir. Geliştirilen bu yöntemleri Frary (1989); Cevaplayıcı Kararları Yöntemleri (Examinee Judgements Methods) ve Doğrudan Cevaplama Yöntemleri (Direct Response Methods) olmak üzere iki kategoride sınıflamıştır.

Cevaplayıcı kararları yöntemlerinde, bireylerden bir maddeye ilişkin bilgi düzeylerine karar vermeleri beklenmektedir. Bunun için bireylerden seçtikleri seçeneğin doğru olduğuna inanma düzeylerini belirtmeleri (güven testi yöntemi) ya da kesinlikle yanlış olduğuna inandıkları seçeneği (alt grup seçme yöntemi) belirtmeleri istenmektedir. Doğrudan cevaplama yöntemlerinde ise bireylerden en doğru olduğunu düşündükleri seçeneği işaretlemeleri beklenmektedir. Bu süreçte bireylerden doğru olduğunu düşündükleri yalnız bir seçeneği işaretlemeleri (seçenek ağırlıklandırma yöntemi); doğru olduğuna inandıkları cevabı, anahtar cevabı buluncaya kadar işaretlemeleri (doğruyu bulana kadar cevaplama yöntemi) ya da farklı sayılarda anahtar cevap içeren ve seçeneklerin her biri aynı zamanda birer doğru-yanlış maddesi olan maddelerin her bir seçeneğini incelemeleri ve doğru olan seçenekleri işaretlemeleri (çoklu doğru cevap yöntemi) beklenmektedir. Bunun yanı sıra Madde Tepki Kuramı'na dayalı olarak da, ağırlıklı puanlanan seçeneklere uygun pek çok model geliştirilmiştir. Bu modellerden bazılarını Ağırlıklandırılmış Cevap Modeli, Kısmi Puan Modeli, Dereceleme Ölçeği Modeli ve Sınıflamalı Tepki Modeli örnek olarak verilebilir.

Alanyazında, ikili puanlamanın bireylerin kısmi bilgilerini yansıtmakta yetersiz kaldığından; ağırlıklı puanlamanın ise bireylerin kısmi bilgilerini daha iyi yansıttığından ve yetenek boyutunun daha büyük bir ranji hakkında daha fazla ve daha güvenilir bilgi sağladığından bahsedilmektedir (Embretson ve Reise, 2000; Ostini ve Neiring, 2006). Yapılan çalışmalar, kullanılan puanlama durumunun istatistiksel analizlerin bulgularını değiştirebildiğini göstermektedir. Selvi ve Alici (2017) yapmış oldukları çalışmada, kullanılan puanlamanın kullanılan istatistiksel yöntemlerin bulgularını anlamlı şekilde değiştirebildiğini bulmuşlardır. Gelin ve Zumbo (2003) yapmış oldukları çalışmada, bir maddenin farklı şekillerde puanlanmasının istatistiksel analizlerin bulgularını değiştirebildiğini bulmuşlardır. Bunun gerekçesi olarak ikili puanlamanın, kullanılan istatistiksel yöntemler için bireyler arasında yeterli değişkenlik oluşturmadığı, bir başka ifadeyle ikili puanlamanın bireylerin ölçülen değişken açısından gerçek durumunu yansıtmakta yetersiz kaldığı gösterilmiştir (Gelin ve Zumbo, 2003; 72). Wetzel, Böhnke, Carstensen, Zeigler ve Ostendorf (2013) yapmış oldukları çalışmada, NEO-PO-R kişilik envanterinin maddelerine ait kategorilerin farklı şekillerde ifade edilip (extreme response style, non-extreme response style) puanlanmasıyla elde edilen veriler üzerinde cinsiyet değişkeni açısından Madde Tepki Kuramı'na dayalı tekniklerle değişen madde fonksiyonu analizleri yapmışlar ve maddelerin puanlanma şekline göre bulgularının değiştiğini bulmuşlardır.

ÖSA açısından ise yapılan alanyazın incelemesi, farklı değişkenler açısından ÖSA'nın işleyişini ortaya koymaya yönelik çalışmaların yapıldığını göstermektedir (Galindo-Garre ve Vermunt, 2006; Güngör, Kormaz ve Sazak, 2015; Güngör Çulha ve Korkmaz, 2011; MacDonald, 2018; Tueler, Drotar ve Lubke, 2011; Uyar, 2015; Wurpts ve Geiser, 2014). Ancak ÖSA'nın farklı puanlama durumlarındaki işleyişini inceleyen bir çalışmaya rastlanmamıştır. Maddelerin puanlama durumlarının ÖSA'dan elde edilen bulguları değiştirebileceği düşünülmektedir. Bu da ÖSA için kullanılacak en uygun puanlama durumunun saptanması ihtiyacını ortaya çıkarmaktadır.

Ulaşılan bulguların alanyazında bu anlamda bulunan eksikliği giderdiği düşünülmektedir. Elde edilen bulgular ÖSA'da en uygun sınıflamaya hangi puanlama durumuyla ulaşılabileceğinin ortaya konması açısından önemlidir.

1.1. Araştırmanın Amacı

Bu çalışmada farklı puanlama durumlarında ÖSA'yla elde edilecek bulguların incelenmesi amaçlanmıştır. Bu amaç doğrultusunda İkili, UYDA (uzman yargısına dayalı ağırlıklı) ve DA (deneysel ağırlıklı) puanlama durumlarında ÖSA'yla ulaşılan en uygun model nasıldır? ve hangi puanlama durumunda ulaşılan model diğerlerinden daha uygundur? sorularına yanıt aranmıştır.

1.2. Araştırmanın Önemi

Çalışma kapsamında ulaşılan bulguların alanyazında bulunan eksikliği gidermesi beklenmektedir. Elde edilen bulgular ÖSA'da en uygun sınıflamaya hangi puanlama durumuyla ulaşılabileceğinin ortaya konması açısından önemlidir.

2. YÖNTEM

2.1. Araştırmanın modeli

Bu çalışmada ÖSA'yla elde edilen bulguların farklı puanlama durumlarında karşılaştırılması amaçlandığından çalışmanın temel araştırma niteliğinde olduğu söylenebilir (Karasar, 1992).

2.2. Araştırmanın çalışma grubu

Dört seçenekli Sözel Akıl Yürütme Yeteneği Testi'ni uygulamak için gerekli örnek genişliği hesaplanırken $n = \frac{t^2 * p * q}{d^2}$ formülünden yararlanılmıştır. Bu formülde,

$$t = \alpha \text{ değeri } 0,025 \text{ için tablo değeri } (t=1,96)$$

$$p * q = \text{Varyans tahmini } (0,5 * 0,5 = 0,25)$$

$$d = \text{Kabul edilebilir hata marjini } (0,05) \text{ (Cochran, 1977)}$$

olmak üzere %95 güvenilirlikle (%5 anlamlılık düzeyi) çalışmada ulaşılmaması gereken kişi sayısı en az 384 olarak hesaplanmıştır. (Bartlett, Kotrlık ve Higgins, 2001). Bu bağlamda çalışma grubunda; 2018-2019 eğitim öğretim yılında Mersin Üniversitesi Eğitim Fakültesi'nde lisans eğitimine devam etmekte olan 595 öğrenci alınmıştır. Çalışmaya katılanların demografik özellikleri Tablo 1'de özetlenmiştir.

Tablo 1.
Çalışma Grubunun Demografik Özellikleri

		n	%
Cinsiyet	Kadın	419	71,0
	Erkek	171	29,0
Mezun olunan lise türü	Anadolu Lisesi	187	40,0
	Düz Lise	244	44,3
	Fen Lisesi	20	3,6
	Anadolu Öğretmen Lisesi	100	18,1
Anne Eğitim Durumu	Okuryazar değil	58	9,9
	İlköğretim	286	48,7
	Lise	184	31,3
	Üniversite	59	10,1
	Lisansüstü	-	-
Baba Eğitim Durumu	Okuryazar değil	29	4,9
	İlköğretim	253	42,8
	Lise	173	29,3
	Üniversite	131	22,2
	Lisansüstü	5	0,8

Çalışmada yer alan öğrencilerin %71'i kadın, %29'u erkek olup düz lise mezunları en yüksek yüzdeye sahip iken (%44,3), fen lisesi mezunları en az yüzdeye (%3,6) sahiptir. Anne ve babaların çoğu ilköğretim mezunu olup, okuryazar olmayanların oranı en düşüktür. Öğrencilerin akademik not ortalaması 75.11, standart sapması ise 10.7'dir.

2.3. Veri toplama aracı

Bu çalışmanın verileri Gözen Çıtak (2007) tarafından geliştirilen ve ikili, DA ve UYDA puanlanabilen 'Sözel Akıl Yürütme Yeteneği Testi' kullanılarak elde edilmiştir. İlgili test 4'er seçenekli, 18 çoktan seçmeli maddeden oluşmaktadır. Test maddelerinin her biri farklı düzeylerde doğruluk içermekte ve '1'-4' puan arasında puanlanabilmektedir. '4' puan maddenin en doğru cevabını, '1' puan ise yine doğru olan ancak en doğru cevaba en uzak ifadeyi içermektedir. İlgili testin geliştirildiği çalışmada testin güvenilirliği, '1-0' puanlama için 0.64 (KR-20), UYDA puanlama için 0.68 (Cronbach α) ve DA puanlama için ise 0.69 (Cronbach α) şeklinde bulunmuştur (Gözen Çıtak, 2007). Buna ek olarak geçerliğine kanıt elde etmek için sayısal ve sözel bölümden öğrenci alan programlara kayıtlı öğrencilerin test puanı ortalamaları arasındaki farkı t testiyle incelenmiş ve her üç puanlama yöntemi için de ortalamalar arasında 0.01 düzeyinde anlamlı fark olduğunu ifade edilmiştir (Gözen Çıtak, 2007). Ayrıca her üç puanlama yöntemi için toplam test puanı ile öğrencilerin Türk Dili Sözlü Anlatım dersi notları arasında hesaplanan korelasyon katsayısının '1-0' puanlama için 0.55; UYDA puanlama için 0.52 ve DA puanlama için ise 0.52 olduğu raporlanmıştır (Gözen Çıtak, 2007).

Çalışma kapsamında maddelerin puanlanması, kullanılan ölçme aracını geliştiren Gözen Çıtak'ın (2017; 33-34) yönergesinde belirttiği şekilde gerçekleştirilmiştir. Buna göre ikili puanlamada maddeler, doğru yanıtlanmışsa '1', yanlış ya da herhangi bir gerekçeyle boş bırakılmışsa '0' puanla puanlanmıştır. UYDA puanlamada on kişilik bir uzman grubunun doğru cevaba yakınlığı oranında her bir maddenin her bir seçeneğine verdiği puanların ortalaması alınarak seçenekler '1-4' puan arasında ağırlıklandırılmıştır. DA puanlamada ise seçenek-toplam çift serili korelasyon katsayılarından yararlanılmış ve elde edilen seçenek ağırlıkları kullanılarak puanlama yapılmıştır. Buna ek olarak UYDA ve DA puanlamada madde, herhangi bir gerekçeyle boş bırakılmışsa '0' puanla puanlanmıştır. İkili puanlama için hesaplanan KR-20 ve ağırlıklı puanlamalar için hesaplanan Cronbach alfa değerleri sırasıyla, 0.60 (ikili), 0.63 (UYDA) ve 0.64 (DA) şeklindedir. Her üç puanlama yöntemiyle elde edilen toplam puanlara ilişkin betimsel değerler Tablo 2'de sunulmuştur.

Tablo 2.

Her Üç Puanlama Yöntemiyle Elde Edilen Toplam Puanlara İlişkin Betimsel Değerler

Betimsel Değerler	Puanlama Yöntemi		
	İkili	UYDA	DA
Aritmetik Ortalama	10,01	58,83	59,13
Medyan	10	59	60
Mod	10	61,00	61,00
Standart Sapma	3,20	6,34	6,17
Çarpıklık	-,09	-,61	-,63
Standart Hata (Çarpıklık)	,10	,10	,10
Basıklık	-,18	,56	,66
Standart Hata (Basıklık)	,20	,20	,20
Minimum Puan	1,00	34,00	33,00
Maksimum Puan	18,00	72,00	72,00

2.4. Verilerin analizi

Çalışma kapsamında elde edilen verilerin analizi ÖSA'yla gerçekleştirilmiştir. ÖSA, Latent Gold 5.1 de yapılmıştır (Vermunt ve Magidson, 2013). ÖSA'yla belirlenen gizli alt gruplardaki bireylerin cevapları birbirine oldukça yakın iken, sınıflar arasındaki farklılıkların da o kadar büyük olması beklenmektedir. Örtük Sınıf Analizi birey odaklı bir istatistiksel yöntem olup, her bir öğrenci kendi cevaplarına göre değerlendirilmektedir. (Lazarsfeld ve Henry,1968) Örtük sınıf analizinde tahminler beklenti maksimizasyonu (Expected Maximization, EM) yapılmaktadır ve en az parametre tahmini ve en az sınıf sayısı ile model veri uyumunu sağlayan en sade modele ulaşılmaya çalışılmaktadır (Moors ve Wennekers, 2003; Silvia, Kaufman ve Pretz, 2009; Collins ve Lanza, 2010). Bu modeli seçmek için ise AIC (Akaike Information Criteria), AIC3 (Akaike Information Criteria-3) ve BIC (Bayesian Information Criteria) gibi bilgi kriterlerinden yararlanılmaktadır (Güngör, Korkmaz ve Sazak, 2015). Çalışma kapsamında AIC, AIC3 ve BIC bilgi kriterlerinden yararlanılarak, tüm kriterler için en küçük değere sahip olan model, en uygun model olarak seçildikten sonra, seçilen modelin sınıflama başarısını değerlendirmek için Entropy R² değerleri hesaplanmıştır. Entropi değeri 0 ile 1 arasında değişmekte olup, 1'e yakın entropi değeri sınıflamadaki yüksek başarıyı göstermektedir (Lukoicene, Varialle ve Vermunt, 2010).

Diğer taraftan ÖSA yerel bağımsızlık varsayımını içermektedir. Yerel bağımsızlık varsayımının sağlanıp sağlanmadığını test etmek için ise her bir sınıf sayısı için ikili artıklar (Bivariate residuals, BVR) değerlendirilmiş olup, maddelerin birbiri ile ilişkileri dikkate alınarak ve bir sınıflı model referans alınarak bilgi kriterleri ve sınıf sayıları elde edilmiştir. Model için hesaplanan bir serbestlik dereceli χ^2 test istatistiğinin 3,84

değerinden daha düşük olması istenmektedir (Vermunt ve Madigson, 2013). Çalışma kapsamında hesaplanan ki-kare değerleri model için hesaplanan bir serbestlik dereceli χ^2 tablo değerinden küçük bulunmuş bu nedenle yerel bağımsızlık sayılısının sağlandığı kanaatine varılmıştır. (İkili puanlama için 0,274, UYDA için 0,266 ve DA için 0,379)

2.5. Araştırmanın etik izinleri

Yapılan bu çalışmada “Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi” kapsamında uyulması belirtilen tüm kurallara uyulmuştur. Yönergenin ikinci bölümü olan “Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler” başlığı altında belirtilen eylemlerden hiçbiri gerçekleştirilmemiştir.

Etik kurul izin bilgileri

Etik değerlendirmeyi yapan kurul adı = Mersin Üniversitesi Sosyal ve Beşeri Bilimler Etik Kurulu

Etik değerlendirme kararının tarihi= 31.05.2019

Etik değerlendirme belgesi sayı numarası=022

3. BULGULAR

Çalışma kapsamında elde edilen bulgular ikili, UYDA ve DA puanlama yöntemleri için ayrı ayrı olmak üzere ikili puanlama için Tablo 3’de; UYDA puanlama için Tablo 4’de; DA puanlama için ise Tablo 5’de sunulmuştur.

Tablo 3.

İkili Puanlama Yöntemi İçin Her Bir Sınıf Modeline Ait Bilgi Kriterleri

Model	Tahminlenen Parametre Serbestlik Derecesi	AIC	BIC	AIC3
1-sınıf	21	574	14653,27	14674,27
2-sınıf	40	555	14313,22	14353,22
3-sınıf	59	536	14088,22	14147,22
4-sınıf	78	517	13994,51	14072,51
5-sınıf	97	498	13980,60	14077,60

Tablo 3 incelendiğinde AIC kriterinin en küçük değere beş sınıflı model için ulaştığı görülürken, BIC kriterinin dört sınıflı model için ulaştığı görülmektedir. ÖSA’da uygun olan modelin seçiminde en az sayıda örtük sınıfla ve en az sayıda parametre tahminiyle açıklanan modele ulaşılmaya çalışılmaktadır. Bu nedenle ÖSA’da sınıf sayısı karşılaştırmalı olarak en az sayıda sınıfla ve en az sayıda parametreyle model veri uyumunu sağlayan model tercih edilerek belirlenmektedir (Moors ve Wennekers, 2003). Bu süreçte kullanılan bilgi kriterlerinin farklı değerler üretmesi durumunda ise, BIC kriterinin örneklem büyüklüğünden daha az etkilenmesi nedeniyle tercih edilmesi önerilmektedir (Güngör, Kormaz ve Sazak, 2015; Lin ve Tai, 2015). Bu bilgiler doğrultusunda ikili puanlama için dört sınıflı modelin en uygun model olduğu kanaatine varılmıştır.

Tablo 4.

UYDA Puanlama Yöntemi İçin Her Bir Sınıf Modeline Ait Bilgi Kriterleri

Model	Tahminlenen Parametre Sayısı	Serbestlik Derecesi	AIC	BIC	AIC3
1-sınıf	54	527	21691,71	21927,41	21745,71
2-sınıf	73	508	21256,98	21575,61	21329,98
3-sınıf	92	489	21101,15	21538,56	21193,15
4-sınıf	111	470	21054,07	21502,71	21165,07
5-sınıf	130	451	21041,01	21608,43	21171,01

Tablo 4 incelendiğinde ikili puanlamada olduğu gibi UYDA puanlamada da AIC kriterinin en küçük değere beş sınıflı model için ulaştığı görülürken, AIC3 ve BIC kriterlerinin dört sınıflı model için ulaştığı görülmektedir. Bu nedenle dört sınıflı modelin UYDA puanlama için en uygun model olduğu kanaatine varılmıştır.

Tablo 5.

DA Puanlama Yöntemi İçin Her Bir Sınıf Modeline Ait Bilgi Kriterleri

Model	Tahminlenen Parametre Serbestlik Derecesi	AIC	BIC	AIC3
1-sınıf	54	527	21691,71	21745,71
2-sınıf	73	508	21275,85	21348,84

Tablo 5 devamı

Model	Tahminlenen Parametre Sayısı	Serbestlik Derecesi	AIC	BIC	AIC3
3-sınıf	82	489	21123,83	21525,38	21215,83
4-sınıf	111	470	21106,35	21590,84	21217,35
5-sınıf	130	451	21093,06	21660,48	21223,06

Tablo 5 incelendiğinde DA puanlama için AIC kriterinin en küçük değere beş sınıflı model için ulaştığı görülürken, AIC3 ve BIC kriterlerinin üç sınıflı model için ulaştığı görülmektedir. Bu nedenle üç sınıflı modelin DA puanlama için en uygun model olduğu kanaatine varılmıştır.

Tablo 3, Tablo 4 ve Tablo 5 birlikte incelendiğinde en az sınıf sayısı ve en az parametre kestirimiyle model uyumuna DA puanlama için ulaşıldığı görülmektedir. Buradan aynı bireyler ve aynı maddeler için ÖSA'nın uygun modeli DA puanlama için sağladığı ifade edilebilir. Buna ek olarak DA puanlama için hesaplanan entropy R^2 değeri 0,74 olarak bulunmuştur. Bu değerden modelin sınıflama başarısının da yeterli düzeyde olduğu söylenebilir (Lukoicine, Varialle ve Vermunt, 2010). ÖSA'da en uygun modele DA puanlamada ulaşıldığından bu puanlama için ÖSA koşullu olasılık değerleri ayrıca hesaplanmış ve her bir madde için Tablo 6'da sunulmuştur.

Tablo 6.

DA Puanlama İçin Örtük Sınıf Analizi Koşullu Olasılık Tahminleri

D1					D2				
Koşullu Olasılıklar	Sınıf	1	2	3	Koşullu Olasılıklar	Sınıf	1	2	3
	Puan					Puan			
1		0,4018	0,5935	0,0047	1		0,6810	0,2063	0,1127
2		0,6829	0,3135	0,0036	2		0,5774	0,4062	0,0165
3		0,7073	0,2314	0,0613	3		0,6930	0,2420	0,0650
4		0,7216	0,1088	0,1696	4		0,7161	0,1179	0,1660
D3					D4				
Koşullu Olasılıklar	Sınıf	1	2	3	Koşullu Olasılıklar	Sınıf	1	2	3
	Puan					Puan			
1		0,6593	0,1998	0,1409	1		0,2233	0,7452	0,0315
2		0,5676	0,3708	0,0617	2		0,5074	0,3893	0,1033
3		0,6969	0,2648	0,0383	3		0,6255	0,3458	0,0287
4		0,7078	0,1383	0,1538	4		0,7531	0,1064	0,1405
D5					D6				
Koşullu Olasılıklar	Sınıf	1	2	3	Koşullu Olasılıklar	Sınıf	1	2	3
	Puan					Puan			
1		0,7058	0,2449	0,0494	1		0,5815	0,4185	0,0000
2		0,7677	0,1592	0,0731	2		0,7062	0,2537	0,0401
3		0,6549	0,2528	0,0922	3		0,7329	0,2007	0,0664
4		0,7012	0,1284	0,1704	4		0,6552	0,0878	0,2570
D7					D8				
Koşullu Olasılıklar	Sınıf	1	2	3	Koşullu Olasılıklar	Sınıf	1	2	3
	Puan					Puan			
1		0,7179	0,2811	0,0010	1		0,3348	0,6650	0,0003
2		0,7561	0,2036	0,0403	2		0,6151	0,3423	0,0426
3		0,7385	0,2066	0,0549	3		0,6493	0,2600	0,0907
4		0,6481	0,1292	0,2227	4		0,7305	0,1226	0,1468
D9					D10				
Koşullu Olasılıklar	Sınıf	1	2	3	Koşullu Olasılıklar	Sınıf	1	2	3
	Puan					Puan			
1		0,5465	0,4532	0,0003	1		0,1531	0,8469	0,0000
2		0,5924	0,3654	0,0422	2		0,2879	0,7121	0,0001
3		0,7451	0,1639	0,0910	3		0,7491	0,1933	0,0576
4		0,6918	0,1124	0,1958	4		0,6991	0,0807	0,2202

Tablo 6 devamı

D11					D12				
Koşullu Olaslıklar	Sınıf	1	2	3	Koşullu Olaslıklar	Sınıf	1	2	3
	Puan					Puan			
1		0,3265	0,6313	0,0422	1		0,3827	0,6173	0,0000
2		0,2654	0,7340	0,0006	2		0,5778	0,4031	0,0191
3		0,7117	0,2126	0,0757	3		0,7147	0,2108	0,0744
4		0,7485	0,0779	0,1737	4		0,7210	0,0783	0,2007
D13					D14				
Koşullu Olaslıklar	Sınıf	1	2	3	Koşullu Olaslıklar	Sınıf	1	2	3
	Puan					Puan			
1		0,6743	0,3104	0,0152	1		0,2796	0,7204	0,0000
2		0,7726	0,2064	0,0209	2		0,7233	0,2746	0,0021
3		0,7209	0,1950	0,0841	3		0,8314	0,1341	0,0345
4		0,6409	0,0850	0,2740	4		0,6053	0,0298	0,3649
D15					D16				
Koşullu Olaslıklar	Sınıf	1	2	3	Koşullu Olaslıklar	Sınıf	1	2	3
	Puan					Puan			
1		0,5539	0,4460	0,0001	1		0,4792	0,5174	0,0034
2		0,4102	0,5843	0,0055	2		0,5439	0,3910	0,0651
3		0,7966	0,1698	0,0335	3		0,6806	0,2552	0,0642
4		0,6408	0,0752	0,2840	4		0,7739	0,0541	0,1720
D17					D18				
Koşullu Olaslıklar	Sınıf	1	2	3	Koşullu Olaslıklar	Sınıf	1	2	3
	Puan					Puan			
1		0,2409	0,7591	0,0000	1		0,7356	0,2600	0,0044
2		0,7039	0,2816	0,0144	2		0,6733	0,3045	0,0222
3		0,7513	0,2149	0,0338	3		0,7342	0,1814	0,0844
4		0,6751	0,0675	0,2575	4		0,6886	0,1385	0,1729
DA Puanlama İçin Örtük Sınıf Olaslıkları									
Sınıf 1	Sınıf 2	Sınıf 3							
0,6970	0,1770	0,1260							

Tablo 6 incelendiğinde öğrencilerin %69,7'si birinci örtük sınıfta (n=405), %17,7'si ikinci örtük sınıfta (n=91) ve %12,6'sı ise üçüncü örtük sınıfta (n=85) yer almaktadır. Sınıf büyüklüğü %69,7 olan birinci sınıfın en yüksek olasılıkla DA puanlamasında üç ve dört puan olarak maddenin en doğru cevabını bulma eğilimini sergileyen örtük grup olduğu gözlenmektedir. İkinci örtük sınıf ise tüm maddeler için çoğunlukla bir ve iki puan olarak maddenin doğru cevabını bulmada zorluk çeken örtük sınıf olmuştur. Üçüncü örtük sınıf ise bazı maddeler için maddenin doğru cevabını yakalayabiliyorken, bazı maddeler için doğru cevaptan uzaklaşan örtük sınıfıdır.

4.TARTIŞMA ve SONUÇ

Yapılan çalışmalar, maddelerin puanlama şeklinin istatistiksel analizlerin bulgularını değiştirebildiğini göstermektedir. (Gulliksen, 1950; Padilla, Hidalgo, Benitez ve Gomez-Benito, 2012; Yurdugül, 2010). Örtük grupların ortaya konmasında kullanılan örtük değişken modellerinden biri olan ÖSA'nın bulgularının da maddelerin puanlama şekline etkilebileceği düşünülmektedir. Bu nedenle ÖSA'da en uygun modele hangi puanlama durumunda ulaşılabileceği konusu merak uyandırmıştır. Bu çalışmada kapsamında da farklı puanlama durumlarında ÖSA'yla elde edilecek bulguların incelenmesi amaçlanmıştır.

Elde edilen bulgular ÖSA'dan elde edilen bulguların farklı puanlama durumlarında değiştiğini göstermektedir. Alanyazında da maddelerin puanlama şeklinin istatistiksel analizlerin bulgularını değiştirebildiğinden bahsedilmektedir. Elde edilen bulgular bu durumu destekler niteliktedir.

Çalışma kapsamında gerçekleştirilen ÖSA'da en az sınıf sayısına en az parametre kestirimiyle DA puanlama için ulaşılmıştır. İkili ve UYDA puanlama için ise ÖSA'da aynı sınıf sayısına ulaşılmış olup elde edilen sınıf sayısı DA puanlamayla ulaşılandan daha yüksektir. Alanyazında ÖSA'da en uygun modelin en az örtük sınıf sayısıyla model veri uyumunu yakalayan model olduğu ifade edilmektedir (Collins ve Lanza, 2010; Silvia, Kaufman ve Pretz, 2009). Bu nedenle en uygun modele DA puanlama yöntemiyle ulaşıldığı ifade edilebilir. ÖSA'da en

uygun modele ikili ya da UYDA puanlama yerine DA puanlamayla ulaşılmamasının nedeni olarak DA puanlamada puanlamanın, veriler elde edildikten sonra sonsal olarak seçenek toplam korelasyonları üzerinden gerçekleştirilmesi ve DA puanlamanın diğer puanlama yöntemlerine nazaran gerek bireylerin bilgi düzeylerini gerekse örtük sınıfların durumlarını daha fazla bilgiyle analize yansıtması gösterilebilir. Elde edilen bulgular Gelin ve Zumbo (2003); Wetzel, Böhnke, Carstensen, Zeigler ve Ostendorf (2013) ve Selvi ve Alıcı (2017)'nin bulgularıyla benzerlik göstermektedir. Buna ek olarak Gelin ve Zumbo (2003); ikili puanlamanın, kullanılan istatistiksel yöntemler için bireyler arasında yeterli değişkenlik oluşturmadığını ve bireylerin ölçülen değişken açısından gerçek durumunu yansıtmakta yetersiz kaldığını ifade etmektedir. İkili puanlamada birey maddeden ancak doğru yanıtı tam olarak bildiği ve işaretlediği durumda puan alabilmektedir. Diğer durumlarda maddeden puan alamamaktadır. UYDA ve DA puanlamada ise birey doğru yanıtı tam bilme ya da kısmi bilme durumlarında da maddeden puan alabilmektedir.

Diğer taraftan çalışma kapsamında elde edilen bulgular İkili puanlama ve UYDA puanlamanın ÖSA'da benzer bulgular ürettiğini göstermiştir. Bu durum ikili puanlamada kısmi bilginin dışlanması, UYDA puanlamada ise puanlamanın önsel olarak uzmanlar tarafından gerçekleştirilmesi nedeniyle uzmanların örtük grupların varlığını bilememeleri ve bu grupların maddeler üzerindeki performanslarını kestirememelerinden kaynaklanmış olabilir. Bu çalışma kapsamında elde edilen bulgular bireylerin kısmi bilgilerinin ve kullanılan farklı puanlama durumlarının ÖSA analizlerini değiştirebileceğine ilişkin önsel kanıt sunmaktadır. Ancak katılımcıların örtük gruplarına ilişkin gerçek durum bilgisi elde olmadığından hangi puanlamanın gerçek duruma en yakın olduğu sonucuna bu çalışma kapsamında ulaşılammıştır. ÖSA'da hangi puanlama yönteminin daha iyi sonuçlar verdiği ortaya konabilmesi amacıyla benzer bir çalışmanın, gerçek grup bilgisinin simüle edilebileceği bir simülasyon verisi üzerinde de gerçekleştirilmesi önerilebilir.

KAYNAKÇA

- Anastasi, A. and Urbina, S. (1997). *Psychological testing*. Prentice Hall/Pearson Education.
- Bartholomew, D.J., Knott, M., Moustaki, I. (2011). *Latent variable models and factor analysis: a unified approach*. London: Wiley.
- Bartlett, J.E., Kotrlik, J.W., and Higgins, C.C. (2001). Organizational research: determining appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal*. 19(1): 43-50.
- Ben-Simon, A., Budescu, D. V., Nevo, B. (1997). A comparative study of measures of partial knowledge in multiple-choice questions. *Applied Psychological Measurement*. 21(1): 65-88.
- Clogg C.C. (1988) *Latent class models for measuring*. In: Langeheine R., Rost J. (eds) *Latent trait and latent class models*. Boston: Springer.
- Cochran, W. G. (1977). *Sampling techniques*. New York: John Wiley ve Sons.
- Collins, L.M., Lanza, S.T. (2010). *Latent class and latent transition analysis: with applications in the social, behavioral and health sciences*. Wiley, New York.
- Embretson, S.E., and Reise, S.P. (2000). *Item response theory for psychologists*. London: Lawrence Erlbaum Associates.
- Frery, R.B. (1989). Partial-credit scoring methods for multiple-choice tests. *Applied Measurement In Education*. 2(1): 79-96.
- Galindo-Garre, F., ve Vermunt, J. (2006). Avoiding boundary estimates in latent class analysis by Bayesian posterior mode estimation. *Behaviormetrika*. 33(1): 43–59.
- Gelin, M.N. and Zumbo, B.D. (2003). Differential item functioning results may change depending on how an item is scored: an illustration with the center for epidemiologic studies depression scale. *Educational and Psychological Measurement*. DOI: 10.1177/0013164402239317.
- Gözen Çıtak, G. (2007). *Klasik test ve madde-tepki kuramlarına göre çoktan seçmeli testlerde farklı puanlama yöntemlerinin karşılaştırılması*. Yayınlanmamış Doktora Tezi, Ankara Üniversitesi, Ankara.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Güngör Culha, D., Korkmaz, M. (2011). Örtük sınıf analizi ile bir örnek uygulama. *Eğitimde ve Psikolojide Ölçme Değerlendirme Dergisi*, 2(2):191-199.
- Güngör, Culha, D., Korkmaz, M., Somer, O. (2013). Çoklu-grup örtük sınıf analizi ve ölçme eşdeğerliği, *Türk Psikoloji Dergisi*. 28, 48-57.
- Güngör, Culha, D., Korkmaz, M., Sazak, H.S. (2015). Örtük sınıf analiziyle yapılan ölçme eşdeğerliği çalışmalarında model seçimi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*. 30(1): 90-105.
- Goddman L.A. (1974) Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*. 61(1): 215-231.
- Hagenaars, J.A. (1990). *Categorical longitudinal data - loglinear analysis of panel, trend and cohort data*. Newbury Park: Sage.
- Karasar, N. (1986). *Bilimsel araştırma yöntemi*. Ankara: Nobel Yayınları.
- Lazarsfeld, P. F., Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.
- Lin, S.W., Tai, W.C. (2015). Latent class analysis of students' mathematics learning strategies and the relationship between learning strategy and mathematical literacy. *Universal Journal of Educational Research*. 3(1):390-395.
- Lukoicene O., Varriale R, Vermunt JK. (2010). The simultaneous decision(s) about the number of lower- and higher-level classes in multilevel latent class analysis, *Social Methodology*, 40(1): 247-283.
- MacDonald, K. (2018). *Latent class analysis*, 18. London Stata Conference . London.
- Madigson, J., Vermunt, J.K. (2004). *Latent class models*. Erişim tarihi: 01.09.2019. (<http://www.statisticalinnovations.com/wp-content/uploads/Magidson2004.pdf>)
- Moors, G., Wennekers, C. (2003). Comparing moral values in western european countries between 1981 and 1999. A multigroup latent-class factor approach. *International Journal of Comparative Sociology*, 44(1):155-172.
- Padilla, J.L., Hidalgo, J.L., Benitez, I., Gomez-Benito, J. (2012). Comparison of three software programs for evaluating DIF by means of the mantel-haenszel procedure; EASY DIF, DIFAS and EZDIF, *Psicologica*. 33 (1): 135-156.
- Rijmen, F. (2011). The latent class model as a measurement model for situational judgment tests. *Psychologica Belgica*, 51, 197-212.
- Selvi, H., Özdemir Alici, D. (2017). Investigating the impact of missing data handling methods on the detection of differential item functioning. *International Journal Of Assessment Tools In Education*. 5(1): 1-14.
- Silva, P.J., Kaufman, J.C., Pretz, J. (2009). Is Creativity domain-specific? latent class models of creative accomplishments and creative self-descriptions. *Psychology of Aesthetics Creativity and the Arts* 3(3)139-148.

- Tueller, S. J., Drotar, S., and Lubke, G. H. (2011). Addressing the problem of switched class labels in latent variable mixture model simulation studies. *Struct. Equ. Modeling*, 18 (1):110–131.
- Uyar, Ş. (2015). *Gözlenen gruplara ve örtük sınıflara göre belirlenen değişen madde fonksiyonunun karşılaştırılması*. Yayınlanmamış Doktora Tezi. Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü. Ankara.
- Ostini, R., Neiring, M.L. (2006). *Polytomous item response theory models*. Sage Publications, Thousand Oaks, California.
- Wetzel, E., Böhnke, J.R., Carstensen, C.H., Zeigler, M., Ostendorf, F. (2013). Do individual response styles matter? Assessing differential item functioning for men and women in the NEO-PI-R. *Journal of Individual Differences*, 34(2), 2013, 69-81. Doi: 10.1027/1614-0001/a000102.
- Vermunt, J.K. (1997). *Log-linear models for event histories*. Thousand Oakes: Sage Publications.
- Vermunt, J.K. ve Magidson, J. (2013). *Technical guide for Latent GOLD 5.0: basic and advanced*. Belmont: Statistical Innovations Inc.
- Wurpts, I.C., Geiser, C. (2014) Is adding more indicators to a latent class analysis beneficial or detrimental? results of a monte-carlo study. *Front. Psychol.* DOI: 10.3389/fpsyg.2014.00920
- Yurdugül, H. (2010). Farklı madde puanlama yöntemlerinin ve farklı test puanlama yöntemlerinin karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*. 1: 1-8.

EXTENDED ABSTRACT

1. Introduction

In the literature, latent variable models are often used to examine whether individuals or variables belong to a latent group or structure (Güngör Çulha ve Korkmaz, 2011). Latent models are classified under four headings as 'factor analysis', 'item response theory', 'latent profile analysis', and 'latent class analysis. Of these, 'latent class analysis was developed by Lazarsfeld (1968) in order to classify individuals by taking advantage of observed variables in dichotomic structure. In the latent class analysis, the probability of the individuals being assigned to the latent classes is calculated statistically using the data obtained from the individuals.

Latent class analysis, in its simplest form, assumes that the population consists of a large number of unobservable groups and formulates individual differences. But it is known that almost all of the latent variable models have strengths and weaknesses. In the literature, many studies have been conducted to reveal the strengths and weaknesses of these models. These studies show that the findings obtained with different latent variable models can be affected by variables such as sample size, item number, test length (Wurpts and Geiser, 2014; Galindo-Garre ve Vermunt, 2006; Tueler, Drotar ve Lubke, 2011).

On the other hand, it is not possible for social sciences to expect equal equivalence of each item and option in psychological space. For this reason, another variable that may change the findings obtained from the LCA is thought to be the scoring method of the items in the measurement tool used.

For this purpose, there are different scoring methods developed in the literature. One of the commonly used scoring methods is dichotomous scoring. Binary scoring (1-0) is often preferred for the scoring of multiple-choice test items because of the ease of scoring it provides. In binary scoring, usually '1' score is given to the correct answers and '0' score is given to the cases other than the correct answer (Gulliksen, 1950).

But binary scoring is criticized on the grounds that it is insufficient to reflect the actual knowledge level of individuals in the literature (Embretson and Reise, 2000). In order to produce solutions to these criticisms, different weighted scoring methods have been developed in which items or options are weighted in different ways. Frary (1989) classifies these developed methods into two categories: Examinee Judgments Methods and Direct Response Methods. The researches indicated that binary scoring was insufficient to reflect the partial knowledge of individuals; weighted scoring reflects the partial knowledge of individuals better and provides more and more reliable information about a larger range of latent dimension (Ostini and Neiring, 2006; Embretson and Reise, 2000). In this study, it is aimed to investigate the findings to be obtained by latent Class Analysis in different scoring situations.

2. Method

The data of this study was obtained by using Verbal Reasoning Test developed by Gözen (2007). Related test consists of 18 multiple choice items with 4 options and the items can be scored as binary, experimental weighted and weighted based on expert judgment. The data were collected from 595 students attending undergraduate education in Mersin University Faculty of Education in 2018-2019 academic year.

The analysis of the data obtained within the scope of the study was performed by latent class analysis. Latent Class Analysis was performed in Latent Gold 5.1 (Vermunt ve Magidson, 2013). While the responses of the individuals in the latent sub-groups determined by latent class analysis are very close to each other, the differences between the classes are expected to be so great. Latent Class Analysis is an individual oriented statistical method and each student is evaluated according to their own answers. (Lazarsfeld and Henry, 1968) In latent class analysis, the predictions are made by Expected Maximization, and the simplest model is tried to be reached with the least parameter estimation and the least number of classes (Moors and Wennekers, 2003; Silvia, Kaufman ve 2003). Pretz, 2009; Collins ve Lanza, 2010). In order to select this model, information criteria such as AIC (Akaike Information Criteria), AIC3 (Akaike Information Criteria-3) and BIC (Bayesian Information Criteria) are used (Güngör, Korkmaz ve Sazak, 2015). Entropy values were calculated in order to evaluate the classification success of the selected model after selecting the most suitable model by using AIC, AIC3 and BIC information criteria. Entropy value varies between 0 and 1, and entropy value close to 1 shows high success in classification.

3. Findings, Discussion and Results

The findings obtained in this study are presented separately as binary, experimental weighted and weighted based on expert judgment scoring methods. Findings showed that the model with four classes is the most suitable

model for binary scoring four-class model is the most suitable model for weighted based on expert judgment scoring and three-class model is the most suitable model for experimental weighted scoring.

When the findings are examined, it is seen that the model fit is reached for experimental weighted scoring with the least number of classes and the least parameter estimation.

It can be stated here that the LCA for the same individuals and the same items provides the appropriate model for experimental weighted scoring. In addition, the entropy R2 value calculated for experimental weighted scoring was found to be 0.74. From this value, it can be said that the classification success of the model is sufficient.

Therefore, it can be asserted that experimental weighted scoring is more consistent with model and less parameter estimation and provides more statistical information due to the fact that both the level of knowledge of individuals and the status of latent classes are reflected in the analysis with more information than other scoring methods. The findings are in line with the findings of similar studies in the literature (Gelin ve Zumbo, 2003; Selvi ve Alıcı, 2017).

The findings of this study provide preliminary evidence that the partial knowledge of individuals and the different scoring situations used may alter the LCA analysis findings. However, since there was no real status information about the latent groups of participants, it was not possible to conclude which scoring was closest to the real situation. A similar study may be proposed on a simulation data in which real group information can be simulated in order to determine which scoring method gives better results.

ÇALIŞMANIN ETİK İZİN

Yapılan bu çalışmada “Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi” kapsamında uyulması belirtilen tüm kurallara uyulmuştur. Yönergenin ikinci bölümü olan “Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler” başlığı altında belirtilen eylemlerden hiçbiri gerçekleştirilmemiştir.

Etik kurul izin bilgileri

Etik değerlendirmeyi yapan kurul adı = Mersin Üniversitesi Sosyal ve Beşeri Bilimler Etik Kurulu

Etik değerlendirme kararının tarihi=31.05.2019

Etik değerlendirme belgesi sayı numarası=022