İKİ ELEŞTİREL AKIL YÜRÜTME GÜCÜ ÖLÇEĞİ İNİN GEÇERLİK VE GÜVENİRLİĞİNİN BİR PSIKOLOGLAR PANELİ TARAFINDAN DEĞERLENDİRİLMESİ*

Richard, B. MODJESKİ
William, B. MİCHAEL
**Çev: Nükhet ÇİKRİKÇİ

ÖZET


Kullanılan 15 standartın herbiri için iki eleştirel akıl yürütmeye testi ile ilgili olarak, her bir panel üyesi ölçegin, standartı karşılama derecesine göre “standart tümüyle karşılıyor belli ölçüde karşılıyor, en alt düzeyde karşılıyor ve hiç karşılımayor” şekilde bir tepkide bulunmuştur. Bu düzeylerden herbirine 3, 2, 4 veya 1 gibi puanlar verilmiştir; ele alınan herhangi bir standart o ölçme aracını uygulanabilir değilse o standart için herhangi bir puan verilmemiştir.

Genel olarak, her iki ölçme aracının standartları karşılama düzeyleri “en alt düzeyde karşılıyor” ile “belirli düzeyde karşılıyor” kategorileri arasında yer almıştır. Bu derecelenemeler ait ortalamalar arasında istatistiksel açıdan manidar bir fark çıkmamıştır.

Bu testlerin yazarlarına araçlarının geçerlilik ve güvenirliklerini artırmak için bir revizyon çalışmasının gerekli olduğunun düşünülmesi önerilmektedir.


**YÖNTEM**

**Örneklem**


**Dereceleme Ölçeği**

Organizasyon ve Biçim: Ele alınan ölçeklerin belirlenen standartlara göre incelemesini sağlamak üzere geliştirilen ilk 19 kişilik gruba gönderilerek anket (12 kişinin yanıtladığı) üç bölümden oluşmuştur: (a) Eğitimsel geçmiş ve halihazırdaaki istihdam durumu, meslekSEL yönelim-
lere ilişkin 6 temel madde içeren 1. bölüm; Cornell'in ve Watson–Glasser'-in ön çalışmada ele alınan 10 temel geçerlik ve 5 temel güvenirlik ölçeme hatası standartını temsil edici, araçların bu standartları karşılama düzeyini ölç克莱meeye izin veren 15 cümleden oluşan 2. ve 3. bölümlerdir. Panelistlere bunları ölç克莱medede 5 seçenek sunulmuştur. Bunlar;

A– Standarti tümüyle karşılar
B– Standarti beli ölçüde karşılar
C– Standart en alt düzeyde karşılar
D– Standarti hiç karşılamayor
E– Standart uygun değil

A, B, C, D, veya E seçeneklerine verilen cevaplar her bir standartla bağlantılı olarak 4, 3, 2, 1 ve 0 şeklinde açıklanılarak puanlanmıştır. Beşinci seçenek ele alınan standartın söz konusu ölçeme aracı için uygulanamaz olması durumunda kullanılabilecek ve puan karşılığı sıfır olan bir seçenektir.

Her bir cümleye karşılık gelen sayısal değer, panel üyesinin görgülü çalışmalardan çıkan Những bulgular ile testin el kitabından sağlanan bilgiler çerçevesinde, her testin bir standart karşılık derecesi hakkında derecelemesini sayısal olarak ifade etme olanağı sağlamıştır.

Temel Standartlar

Bu çalışmada ele alınan testleri inceledede yararlanılan standartlar; geçerli alanında temel olan 10 standart ile güvenirlik ve ölçme hatası konusunda temel olan 5 standarttır. Tercih edilir, edilmeme şeklinde sınıflandırılan diğer standartlar inceledede göz önünde alınmamıştır. E bölümünde ele alınan geçerliğe ilişkin standartlar, temel olarak sınıflandırıldıkları her durumda $E_1, E_2, E_3, E_4, E_5, E_6, E_7, E_8, E_9, E_{10}, E_{12}, E_{13}$ şeklinde kodlanmıştır. Buna karşılık güvenirlik ve ölçme hatasıyla ilgili standartlar F bölümünde ele alınmış ve temel olarak sınıflandırıldıkları her durumda $F_1, F_2, F_3, F_4$, ve $F_6$ sembolleriyile gösterilmiştir.

Geçerlik Standartları

*Genel İlkeler*

$E_1$ bir el kitabı veya araştırma raporu testin kullanılmasına önerildiği her çekarsama türü için geçerliğe ilişkin kantları içermelidir. Eğer geçerlik, öne sürülene en uygun yorumlamalar için araştırılmış ve incelenmemişse bu durum açıkça kavuşturulmalıdır.
E₂ Bir test uygulayıcısı, geçerlik ve güvenirlik iddialarını destekleyecek kantlar sunmakla yükümlüdür. Karara ulaşmada test puanlarının kullanım, yeterli derecede канıta desteklenmiş olmalıdır.

**Ölçüt Dayanaktı Geçerlik**

E₃ Bütün ölçüt ölçüleri tam ve açık bir şekilde testin el kitabında belirtilmelidir. El kitabı veya araştırma raporu ölçünün uygunluğunu konusundaki yürütülmemelədən bulunmamalıdır. Olabildiğince, performansın, ölçüt ölçüsünün yansıttığı yönlerine ve onu etkileyecek ilişkisini etmenlere dikkat çekmelidir.

E₄ Bir geçerlik saptama çalışmasında, seçilen örneklem ve testin uygulandığı koşullar, testin kullanım için önerilen koşullarla tutarlı ve okuyucunun kendi durumunu ile ilişkisi konusunda karara varabilmek için yeterince açıklanmamalı ya da betimlenmelidir.

E₅ Bir geçerlik çalışması için verilerin toplanmasında, çalışmanın amaçları ile tutarlı işlemleri izlenmelidir.

E₆ Birey hakkında yapılan tahmin ve yargılamalara ne kadar güvenle bakılabileceği konusunda okuyucuya yardımcı olacak “ölçüt dayanaktı geçerlik” verilerine ilişkin her türlü istatistiksel analize teslim el kitabında yer verilmelidir.

E₇ Bir test kullanıcısı testteki veya test maddelerindeki yanlışlıkları incelemelidir. Testi kullanan kişi için, mümkün olduğu etnik, einsiyet ve diğer alt gruplar (kültürl) açısından, ölçüt dayanaktı geçerlıkte ortaya çıkmış olası farklılıkların araştırılması bu gruplara ilişkin verilerin el kitabında bulunması, açıklanması gerekir.

E₈ Bir puanlama anahtarına sahip ölçme araçının maddelerinin seçimi veya testin (maddelerinin) ağırlıklandırılması, yalnızca bir tek örnekleme dayah olarak yapıldığında, el kitabının bağımsız örneklemlerden elde edilen verilere dayah olarak, bir veya daha fazla sayıdaki çapraz geçerlik (Cross-validation) katsayılarının içermesi gerekir. Geçerlik ifadeleri orijinal örnekleme bağlı olarak verilmemelidir.

**Kapsam Geçerliği**

E₁₂ Eğer testteki performans, evrendeki performansı temsil eden bir örneklem olarak yorumlanıyorsa; testin el kitabı, evreni ve evrenden alınan örneklemi seçmede takip edilmiş olan işlemleri açık bir biçimde tanımlamalıdır.
**Yapı Geçerliği**

E₁₃ Eğer yazar, bir testten alınan puanları teorik bir değişkeni (yetenek, tutum v.b. psikolojik özellikleri) ölçüyor olarak yorumlamayi önerirse, önerdiği yorumu tam olarak yer verilmelidir. Yazarın öne sürdüğü kuramsal yapılı, diğer kuramlar temelinde ortaya çıkan yorumlamaldardan ayırdıdebilmelidir.

**Güvenirlik ve Ölçme Hatası Standartları**

**Genel İlkeler**

F₁ Testin el kitabı veya araştırma raporu testin amaçlanan kullananın açısından puanların yeterli olup olmadığı kullanım yanılışlaması olanak veren güvenirlik hesaplamalarını içermelidir.

F₂ Güvenirlik katsayısını ya da ölçmenin standart hatası belirledede kullanım işlem ve örneklemeler, test uygulayıcısının, rapor edilen verileri ilgilendiği birey veya gruplara uygulanabilir nitelikte olup olmadıklarına karar verebilmesine olanak sağlamalıdır.

F₃ Güvenirlik çalışmaları dair raporlar, sırasıyla hata bileşenlerinin varyansları ölçmenin standart hatası da çarpım momentler korelasyon katsayısını içinden el kitabında ifade edilmelidir. Verilerin iyi bilinmeyen bir biçimde ifade edilmesi olmalı durumunda, gelişimlerine göndermeler yapılabacak açık bir biçimde betimlenmesi gerekir.

**Formların Karşılaştırmalıbilirliği**

F₄ Eğer bir testin aynı cevaplayıcı grubu için kullanılmak üzere geliştirilmiş iki veya daha fazla formu varsa, formlardaki ortalama, varyans ve medde özellikleri hakkında bilgiler, bu formlardan alınan puanlar arasındaki korelasyon katsayılarıyla birlikte el kitabında raporlaştırmalıdır. Eğer bu türden kanıtlar elde edilmemisse, testin el kitabı okuyucuyu formların denklığını kabul etme konusunda uyarmalıdır.

**Formların Uygulamalararası Sonuçların Karşılaştırmılması**

F₆ Testin el kitabı puanlarının ne ölçüde kararlı olduğunu yani, belirli bir zaman aralığıyla bir testin paralel formu uygulanmışsa, puanların ne kadar kararlılık gösterdiği belirtmelidir. Ayrıca testin kullanışlılığındaki boyolesi herhangi bir varyasyonun etkisi de el kitabında açıklanmalıdır. İki uygulama arasında geçen sürenin uzunluğu testin doğasına ve test puanlarının nasıl yorumlanacağına bağlıdır.
İstatistiksel Analizler

Verilen bir standart için alınan cevap kategorilerine ilişkin olarak yapılan derecelmelerin sıklığı her ölçme aracının bulunmuş ve raporlaştırılmıştır. Panel üyelerinden tayin edilen derecelmelerin ortalaması, medyanı, ranji ve standart kayması her bir standart için belirlenmiştir. Buna ek olarak hariki ölçme aracı için geçerlik, güvenirlık standartlarını karşılama düzeyleri bakımından yapılan ortalama derecelerin ortalamaları bulunmaktadır.

Ayrıca panel üyelerinin herbir geçerlik, güvenirlık ve ölçme hataları standartı için tayin ettikleri ortalama derecelemelerin puvanlarından ortalamaları arasındaki farkın manidarlığı yine her iki araç için test edildi. Son olarak panel üyelerinin 14 standart için tayin ettikleri derecelendirme melerin tümüne ilişkin kategori iç derecelmelerin hesabı toplam veya ortalama olarak elde edildi (Guiford, 1965, s. 300).

BULGULAR

Panel üyelerince yapılan derecelmelerle ait veriler her iki ölçekte iki verilen: a) yargıcılararası güvenirlilik tahminleri açısından b) çoğunluğu betimleyici olmak üzere, panel üyeleri tarafından iki araçın 10 geçerlik standartını karşılama derecesine ilişkin olarak derecelmeler açısından c) iki araçın beş güvenirlık ve ölçme standartlarını karşılama derecesine ilişkin olarak yapılan derecelmelerinin paralel betimsel istatistikleri aşağıda verilmiştir.

1. CET ve WG-EAGÖ için sırasıyla 11 ve 12 psikolog tarafından, ortalama derecelmelerle dayanıklar hesaplanan "yargıcılararası güvenirlik katsayısı (rkk)" .84 ve .88 (— .001) olarak hesaplanmıştır.

2. Panel üyelerinin Cornell ve Watson-Claser için 10 geçerlik standartını karşılama düzeylerine ilişkin olarak verdikleri derecelmeleri için aşağıdaki istatistiksel sonuçlar Tablo 1'de gösterilmiştir.

a) Beş derecelemeye kategorisi (A, B, C, D veya F) üzerinden yapılan derecelmelerin dahil olduğu önemli runlar her iki ölçümde vardı ve her standarda ilişkin olarak hem her kategori en az bir frekans ahyordu. Örneğin F2 standartı, her kategoriye en az bir defa, söz konusu araç için konuşmuştur.

b) Eğer bir geçerlik standartı "olumlu (tercih edilebilir) derecelendirme rebilmek için panel üyelerinden yüzde 50'sinden ya da daha fazlasından
ya Standardı tümüyle karşılıklar (kategori A) ya da “Standardı kısmen karşılıklar (kategori B)” derecelendirmesi almaktadır. “ölçütü konulursa”Cornelliçin iki standardın (F₆ ve F₇), Watson- Glaser için beş standardın (E₁, E₂, E₆, E₇ ve E₁₂) ölçütü karşılığı gözlenmiştir.

c) Her iki ölçme aracı da E₉ ve E₁₀ standartları açısından, en az sikkhla A ve B kategorilerinde yer almıştır.

d) 10 geçerlilik standardına ilişkin olarak yapılan derecelenmeler üstünden hesaplanan ortalama ve standart sapmalar Cornell için, 2.22 ve .42; Watson-Glaser için, 2.31 ve .48'dir.

Bu işlemde, bir derecelendiricinin (panel üyesi) 10 geçerlilik standardı açısından, iki ölçme aracı için de yaptığı iki derecelene takdiminin birbirinden bağımsız olduğu sayılmasını vardır. Bu sayıyla, bir derecelendiricinin belli bir standarda karşı oluunuoluamsız yanılışlarının bir çevap kurulumu (response set) yaratması olasılığı açısından sorgulanmıştır. Bu amaçla hesaplanan ortalamalar arasındaki .09'lu fark istatistiksel olarak manidar bulunmamıştır (t (18)=.43, P < .05).

3. Cornell ve Watson-Glaser güvenilirlik ve ölçme hatasına ilişkin beş standardı karşılama düzeylerine ilişkin olarak panel üyesinin yapıtları derecelenmelerden çıkarın ve aşağıda verilen istatistiksel sonuçlar Tablo 1'de özetlenmiştir.

a) Geçerlilikteki gibi, güvenilirlikte, beş kategoride verilen cevapların sıklığına ilişkin yapılan derecelenmeler, geçerlilık ki derecelenmelerden daha yüksektir.

b) Eğer aynı standart sözkonusu araç için cevaplayıcıların yüzde50'şi si tarafından A ve B derecelene kategorisine konmuşsa, verilen standart ifadesinin, güvenilirlik ve ölçme hatası standartına uygunluğunu temsil ettiği kabul edilmiştir. Bu bulgu, Cornell için F₂ ve F₃ standartlarında; Watson-Glaser için F₁,F₂,F₃ ve F₄ standartlarında sağlanmıştır.

c) Her iki ölçme aracı F₆ standardı açısından A kategorisinde bir derece almamış B kategorisinde de sadece bir derecelene elde etmiştir.

d) Sözkonusu güvenilirlik ve ölçme hatası standartları açısından yapılan beş derecelenmenin ortalama ve standart sapması sırasıyla, Cornell için, 2.46 ve .74; Watson-Glaser için, 2.74 ve .79 bulunmuştur.

Yine iki ölçme aracı için yapılan beşli iki takım derecelene işleminin birbirinden bağımsız olduğunu saytıslma test etmek için, bu derecelenmelere ait ortalamalar arasındaki .28'lik fark istatistiksel olarak incelenmiş ve manidar bulunmamıştır (t (18)=.42, P < .05).
<table>
<thead>
<tr>
<th>Test</th>
<th>Verilen Derece</th>
<th>Geçerlik İfadeleri</th>
<th>Güvenirlik İfadeleri</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>E1</td>
<td>E2</td>
</tr>
<tr>
<td>Cornell A. Standart tümüyle karşılıyor</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>B. Standart kısmen karşılıyor</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>C. Standart en alt düzeyde karşılıyor</td>
<td>4</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>D. Standart hiçbir karşılıymaz</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>E. Standart Uygulanamaz</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Aritmetik Ortalama</td>
<td>2.4</td>
<td>2.5</td>
<td>2.1</td>
</tr>
<tr>
<td>Ortanca</td>
<td>2.4</td>
<td>2.4</td>
<td>2.0</td>
</tr>
<tr>
<td>Dizi Genişliği (Ranj)</td>
<td>1-4</td>
<td>1-4</td>
<td>1-4</td>
</tr>
<tr>
<td>Watson A. Standart tümüyle karşılıyor</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>B. Standart kısmen karşılıyor</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>C. Standart en alt düzeyde karşılıyor</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>D. Standart hiçbir karşılıymaz</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>E. Standart Uygulanamaz</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Aritmetik ortalama</td>
<td>2.3</td>
<td>2.4</td>
<td>2.0</td>
</tr>
<tr>
<td>Ortanca</td>
<td>2.5</td>
<td>2.5</td>
<td>2.0</td>
</tr>
<tr>
<td>Dizi Genişliği (Ranj)</td>
<td>1-4</td>
<td>1-4</td>
<td>1-4</td>
</tr>
</tbody>
</table>

Not: A, B, C, ve D kategorileri 4, 3, 2, ve 1 puan almış E kategorisine puan verilmemiştir.
a Geçerlik $\bar{X}_g = 2.22$ $\sigma_g = .42$ Güvenirlik ve Ölçme Hataları $\bar{X}_g = 2.46$ $\sigma_g = .74$
bGeçerlik $\bar{X}_g = 2.31$ $\sigma_g = .48$ Güvenirlik ve Ölçme Hataları $\bar{X}_g = 2.74$ $\sigma_g = .79$
SONUÇLAR

Geçerlık ve güvendiğlik standartlarından yola çıkılarak bir grup uzmanın bu çalışmada yaptığı derecelenmelerin yorumu şu şekildedir:

1. Geçerlik konusunda ileri sürülen genellemeler,

a) Geçerlikte, “temel” olarak sınıflandırılan standartlar açısından söz konusu ölçme araçlarının bu standartları karşılama düzeylerindeki, panel üyesi psikologların algılamaları arasında bir değişim gözlemlenmiştir.

   b) Her iki araç, nispeten geçerlik hesaplamalarında başvurulan iki örneklem hakkında gerekli bilgileri vermeleri bakımından (E₆ standartı) ve höyresi bir amaç için veri toplam koşulların gözönüne alındığında, olduzkça yüksek puan düzeyine sahip cevap kategorilerinde derecelendirilmiştir.

Her iki ölçme aracı da olası test yanılığı bağımlılındaki değerlendirilmeleri ve çapraz geçerlik çalışmalardaki eksikliği karşılama bakımından oldukça yüksek derecelenmeler elde ettiler.

Watson-Glaser, tek bir standart açısından (evreni yansıtıç davranışları örneklemleyen maddelerin seçiminde takip edilen işlemlerin tanımlanması ile evrenin tanımlanması standartı) E₁₂ Cornel’le göre daha üstün olarak görülüp, derecelendirilmişdir.

c) Her iki araç için 10 geçerlik standardına ilişkin olarak, panel üyesinin tayin ettiği derecelerin ortalamaları arasında manidar bir fark gözlenmemesine rağmen, Watson-Glaser temel geçerlik standartları açısından Cornell’e göre daha üstün bir araç olarak görülmuş, panel üyesinin en az % 50’si tarafından bu araçlar A ve B kategorilerine (Watson-Glaser dört standart açısından, Cornell ise iki standart açısından) atanmıştır.

2. Güvendiğlik ve Ölçme Hatasa İlişkin Olarak İleri Sürülen Genellemeler

a) Geçerlikte olduğu gibi, bunda da her iki ölçme aracının temel standardlardaki beş standardı karşılama düzeylerine ilişkin olarak panel üyesinin derecelenmeleri arasında bir değişim çıkmıştır.

b) Her iki araç da (i) “güvendiğlik katsaylarının ve ölçme hatalarının belirlenmesinde kullanılan örneklemnin seçimi ve takip edilen işlemlerin tanımlanmasına” (F₂ standartı); (ii) “Güvendiğlikle ilgili verilerin raporlaştırılmasında kabul edilen yöntemlerin kullanılması bakımından” (F₃ standartı), tercih edilir araçlar olarak belirlenmiştir.
Cornell ve Watson-Glaser, her ikisi de “paralel formların kullanılması bağlamında, puanların kararlılığının raporlaştırılması içeren standartta ve bu paralel formların uygulamaları arasında geçen süre ile ilgili standartta” (F₆ standartı) negatif üçta derecelmeler elde ettiler.

Aşağıdaki standartlar açısından Watson-Glaser, Cornell’e göre üstünlük sergilemiştir:

Watson-Glaser, Cornell’e göre (i) ölçme hatası ve güvenirlik ile ilgili araştırma raporları ve el kitabında bununla ilgili kanıtlar bulunması (F₁ standartı) (ii) testin iki veya daha fazla formunun birbirine denkliği konusundaki verilere sahip olması (F₄ standartı) açısından daha üstün bir araç olarak görülmüştür. (Cornell de eleştirel akıl yürütmenin karşılaştırılabilir düzeylerini yansıtabacak iki formun düzenlenmesinin eksikliği, muhtemelen bu araçın tereci edilmeyecek bir şekilde derecelenmesine yol açmıştır). (iii) Panel üyelerinin ölçme hatası ve güvenirlik standartları açısından yaptıkları derecelmelerle ait ortalama puanlar arasında manidar fark olmamasına rağmen, Watson-Glaser temel standartları karşılaması bakımından daha üstün bir araç olarak görülmüştür. Elde edilen derecelmelerin herhangi biri, panel üyelerinin en az % 50’si tarafından “standardı tümüyle karşılıyor” (A kategorisi) ve “standardı belli ölçüde karşılıyor” (B kategorisi) kategorisinde, sırasıyla, Watson-Glaser dört standartta, Cornell iki standartta derecelmeler elde etmiştir.

ÖNERİ

Eğer standartlardan sağlanan ölçütlер, iki aracın değerlendirilmesi için uygunsa, güvenirlik, geçerlik belirlemesine ilişkin araştırma ve geliştirme çabaları ortaya çıkabilecektir. Gelecekte her iki aracın da gözden geçirilmesi düşünülmelidir.

Kaynaklar


