

Performans Değerlendirmelerinde Geleneksel Olmayan Değerlendirici Kaynaklarının Kullanımı

H. Canan Sümer*

Reyhan Bilgiç

Orta Doğu Teknik Üniversitesi

Özet

Bu çalışmanın amacı, bir kamu kuruluşunda çok kaynaklı performans değerlendirmesinin uygulanabilirliğini araştırmak ve geleneksel olmayan değerlendirici kaynakları (astlar, çalışma arkadaşları ve öz) tarafından yapılan performans değerlendirmelerini hem psikometrik açıdan hem de kullanıcı tepkileri temelinde incelemektir. Özellikle ast değerlendirmeleri için kaynak içi değerlendirmeler arasındaki korelasyonlar (değerlendiriciler arası tutarlılık katsayısı) (.49) oldukça yüksek bulunmuştur. Kaynaklar arası korelasyonların örüntüsü ise yazında rapor edilen örüntülere paralellik göstermiştir (örn., yönetici-çalışma arkadaşları = .31, öz-çalışma arkadaşları = .30, ast-yönetici = .26, öz-yönetici = .17). Yapılan varyans analizleri ve izleyen ikili karşılaştırmalar sonunda, öz ve çalışma arkadaşı değerlendirmelerinin cömertlik etkisine en yatkın değerlendirmeler olduğu, amir değerlendirmelerinin ise cömertlik etkisine görece daha az yatkın değerlendirmeler olduğu bulunmuştur. Bununla birlikte, hale etkisi en az öz değerlendirmelerinde gözlenmiştir. Geleneksel olmayan kaynakların kullanımına yönelik kullanıcı tepkilerinin yorumuyla birlikte, çok kaynaklı performans değerlendirmesi yaklaşımının Türkiye bağlamında uygulanabilirliği tartışılmıştır.

Anahtar kelimeler: Çok kaynaklı değerlendirme, cömertlik etkisi, hale etkisi, kullanıcı tepkileri

Abstract

The purpose of this study was to evaluate the feasibility of a multi-source feedback system in a public sector organization and, more specifically, to examine performance appraisals provided by nontraditional rating sources (i.e., subordinates, peers, and self) both psychometrically and on the basis of user-reactions. Intraclass correlations (i.e., inter-rater reliabilities) for subordinate evaluations were found to be relatively high (.49). The pattern of correlations between ratings provided by different source pairs were similar to those reported in the literature (e.g., supervisor-peer = .31, self-peer = .30, subordinate-supervisor = .26, self-supervisor = .17). Results of a series of ANOVAs indicated that while self and peer ratings were most prone to leniency, supervisory ratings were relatively less prone to this bias. Self ratings were less open to halo than the ratings provided by the other sources. Along with an evaluation of the obtained user reactions, applicability of a multi-source performance appraisal system within the Turkish context is discussed.

Key words: Multi-source evaluations, leniency, halo, user reactions

*Yazışma Adresi: Doç. Dr. H. Canan Sümer, Orta Doğu Teknik Üniversitesi, Psikoloji Bölümü 06531, Ankara.

E-posta: hcanan@metu.edu.tr

Yazar Notu: Bu makalede ifade edilen görüşler, çalışmanın yürütüldüğü kurumun görüşlerini değil, yazarların görüşlerini temsil etmektedir.

Geleneksel amir/yönetici değerlendirmelerine alternatif bir yaklaşım olarak geliştirilen çok kaynaklı geribildirim yaklaşımı, performans değerlendirmesi alanında son 10-15 yılda yaşanan en belirgin yapısal değişikliktir (Levy ve Williams, 2004). Yazında 360 derece geribildirim yaklaşımı olarak da bilinen, çok kaynaklı değerlendirme yaklaşımının temelinde, birden çok kaynağın (örn., astlar, kişinin kendisi/öz, çalışma arkadaşları, yönetici ve müşteriler gibi) sağlayacağı bilginin, tek bir kaynağın sağlayacağı bilgiden daha kapsamlı ve geçerli olacağı varsayımı bulunmaktadır (Antonioni, 1996; Bracken, Timmreck, Fleenor ve Summers, 2001; Dalessio, 1998; London ve Smither, 1995). 1990'lı yılların ortalarında Amerika Birleşik Devletlerinde şirketlerin %25'inin çok kaynaklı geribildirim yaklaşımını kısmen ya da tamamen kullandıkları tahmin edilmekteydi (Antonioni, 1996). Bu oranın günümüzde artmış olması büyük bir olasılıktır.

Bu yaklaşımın popülaritesinin artmasında (bkz., Dalessio, 1998; London ve Beatty, 1993; London ve Smither, 1995; Westerman ve Rosse, 1997), daha katılımcı bir performans yönetimi anlayışını temsil etmesi, çalışanların etiketlenilmesi ya da sınıflandırılmasından çok geliştirilmesini hedeflemesi, performans değerlendirmelerini bir kişinin sübjektif değerlendirmesine bağımlı kalmaktan kurtarması gibi özelliklerinin önemli rol oynadığı söylenebilir. Bu bağlamda çok kaynaklı değerlendirme; takım çalışması olmadığı durumlarda bile koordineli çalışmanın önem kazandığı, işler arası sınırların hızla eridiği iş ortamlarında, hızlı değişimlere ayak uydurabilmek için yetkinliklerini sürekli geliştirmesi, inisiyatif alması, karar vermesi ve problem çözmesi gereken günümüz çalışanın beklentileriyle daha uyumlu bir performans yönetimi yaklaşımıdır. İşin doğasındaki değişimlere paralel olarak, çalışanların demografik kompozisyonlarındaki değişimler de, çok kaynaklı değerlendirmeyi kolaylaştıran önemli bir diğer faktör olmuştur. İşlerin giderek daha karmaşık hale gelmesi, çalışanlarda aranan niteliklerin de değişmesine neden olmuştur. Karar verme, problem çöz-

me, analitik düşünme ve muhakeme gibi yetenekler, birçok işte aranan temel özellikler arasına girmiştir. Sonuç olarak, bahsedilen özelliklere sahip çalışanların, performans değerlendirme sürecine katkıda bulunma potansiyeli ve de talebi ortaya çıkmıştır. Kısaca, işin doğasında ve çalışanlarda gözlenen değişim ve dönüşümler, performans değerlendirmesine çok kaynaklı yaklaşımı destekleyici ve teşvik edici bir ortamın oluşmasını sağlamıştır.

Çok kaynaklı geribildirim sistemlerinde yapılan değerlendirmeler, çalışmaları geliştirmek amacıyla yapılmaz, terfi, ücretlendirme ve ödüllendirme gibi daha yönetsel kararlara hizmet edecek şekilde de kullanılmaktadır (Borman, 1997). Antonioni (1996), çok kaynaklı değerlendirme sistemlerinin geribildirim/gelişim dışındaki amaçlara yönelik kullanımındaki artışa dikkat çekmektedir. Bu gelişmeye rağmen, çok kaynaklı geribildirim sistemlerinin temel amacı çalışana, performansını iyileştirici, yapıcı geribildirim sağlamak, çalışanın gelişimine katkıda bulunmaktır (bkz., Garavan, Morley ve Flynn, 1997). Nitekim, geribildirim ve bireyi geliştirme amacıyla yapılan çok kaynaklı değerlendirmelerin kalitesinin, yönetsel amaçlı yapılan çok kaynaklı değerlendirmelerin kalitesinden daha yüksek olduğu da görgül çalışmalarla gösterilmiştir (bkz., Dalessio, 1998). Ayrıca, gelişim amaçlı çok kaynaklı değerlendirme sistemlerinin, yönetsel amaçlı sistemlere oranla daha kolay işletildiği ve kullanıcılar tarafından daha çok kabul gördüğü de söylenebilir.

Çok kaynaklı değerlendirme sistemlerinde, farklı kaynaklardan gelen ve mümkün olan durumlarda değerlendirmeyi yapan değerlendiricilerin (yönetici dışında) kimliklerinin gizli tutulduğu değerlendirmeler (Antonioni, 1996), kaynak ortalamasını yansıtabilecek şekilde hem betimsel hem de görsel olarak değerlendirilen kişiye iletilmektedir. Bazı durumlarda, farklı kaynakların yaptıkları değerlendirmelerin geribildirimini sadece yazılı olarak yapılırken, değerlendirmeler genellikle sözlü bir geribildirim seansında, amir ya da bir danışman

("facilitator" veya "coach") tarafından kişiye iletilmektedir (Seifert, Yukl ve McDonald, 2003).

Çalışmalar, farklı değerlendirici kaynaklarının, performansın farklı yönlerine yönelik gözlem yapma şansına sahip olduklarını göstermektedir. Yöneticilerin özellikle, çalışanın görev performansını değerlendirmede etkin ve yetkin olduğu bilinmektedir. Çok kaynaklı değerlendirme yaklaşımından bağımsız olarak, yönetici değerlendirmeleri halen en yaygın olarak kullanılan değerlendirme kaynağıdır (Murphy ve Cleveland, 1995). Çalışma arkadaşları ya da denker, performansın, yöneticiler tarafından gözlenemeyen yönlerini gözleme şansına sahiptir. Çalışma arkadaşlarının görev performansı ile birlikte, ortamsal performansa yönelik değerlendirme konusunda etkin olacağı belirtilmektedir (bkz., Murphy ve Cleveland, 1995). Amirleri tarafından performanslarının gözlendiğini bilen çalışanlar, iyi değerlendirme alma çabasıyla davranışlarını değiştirebilirken, çalışma arkadaşları daha gerçekçi performans gözlemleri yapma şansına sahiptir.

360 derece geribildirim yaklaşımının dışında, ast değerlendirmesi çok yaygın olmayan bir uygulamadır. Ancak, ast değerlendirmelerinin işlevselliğini gösteren görgül çalışmalar bulunmaktadır (örn., Atwater, Roush ve Fischtal, 1995). Astların da özellikle, kendi yöneticilerinin ortamsal performansına yönelik bilgiye sahip oldukları söylenebilir. Astların ve çalışma arkadaşlarının değerlendirmesini, yönetici değerlendirmelerine kıyasla daha etkin kılan önemli bir özellik de birden fazla kişinin yaptığı değerlendirmelerin birleştirilmesidir.

Çok kaynaklı değerlendirme yaklaşımlarının bir diğer temel unsuru öz değerlendirmeleridir. Çalışanların kendi performanslarını daha cömertçe değerlendirme eğiliminde olduğu ve bu değerlendirmelerin yönetici değerlendirmeleriyle yüksek korelasyon göstermediği bilinmektedir (örn., Thornton, 1980). Çok kaynaklı değerlendirme sistemleri içinde öz değerlendirmelerinin amacı, kişinin performansına yönelik algılarıyla, diğer kay-

nakların kişi için yaptıkları değerlendirmeler arasındaki farka dikkat çekerek, değişim ve gelişim için bir farkındalık ve motivasyon yaratmaktır (London ve Smither, 1995).

Farklı değerlendirici gruplarının kişinin performansının farklı yönlerini gözleme ve değerlendirme şansına sahip olduğu görüşünden yola çıkarak, değerlendirici gruplarının yaptıkları değerlendirmeler arasında çok yüksek pozitif korelasyonlar beklenmemektedir. Kaynaklar arası korelasyonları inceleyen çalışmalar genel olarak bu beklentiye destekler niteliktedir. Örneğin, Harris ve Schaubroeck'ın (1988) farklı değerlendirme kaynaklarına yönelik meta-analiz çalışmasında; yönetici, çalışma arkadaşı ve öz değerlendirmeleri karşılaştırılmıştır. Bu meta-analizin sonuçlarına göre, çalışma arkadaşları-yönetici değerlendirmeleri arasındaki ortalama korelasyon .62 iken, öz-yönetici korelasyonu .35, öz-çalışma arkadaşı korelasyonu ise .36'dır. Bir diğer meta analiz sonunda Conway ve Huffcutt (1997), yapılan değerlendirmelerin kaynak içi ve kaynaklar arası korelasyonlarını incelemiştir. Bu çalışmanın sonunda, düzeltilmiş ortalama korelasyonlar, yönetici-çalışma arkadaşı değerlendirmeleri için .34, ast-yönetici için .22, ast-çalışma arkadaşı için .22, ast-öz için .14, öz-yönetici için .22, öz-çalışma arkadaşı için ise .19 olarak tespit edilmiştir. Kaynak içi tutarlılık katsayıları ise, yöneticiler için .50, çalışma arkadaşları için .37, astlar için ise .30 olarak bulunmuştur. Daha yakın bir zamanda yapılan bir çalışmada ise farklı kaynakların yaptıkları değerlendirmeler arasındaki korelasyonların çalışma arkadaşı-yönetici arasında .30, öz-yönetici arasında .15 ve ast-çalışma arkadaşı arasında -.05 olduğu tespit edilmiştir (Atkins ve Wood, 2002).

Diğer birçok yönetim kuram, model ve yaklaşımlarında olduğu gibi, Kuzey Amerika kökenli bu yaklaşım, ülkemizin de dahil olduğu birçok diğer ülkede ilgi ve heyecanla karşılanmıştır. Örneğin, Türkiye'de farklı sektörlerde hizmet veren 200 özel şirkette insan kaynakları yönetimi faaliyetlerinin kapsamı üzerine yapılan bir çalışmada, çok kay-

naklı değerlendirme sistemlerinin bütün olarak olmasa da bazı unsurlarıyla kullanımının küçümsenmeyecek bir düzeyde olduğu görülmüştür (Sözer, 2004). Ancak insan kaynakları yönetimi uygulamalarının kültürle uyumlu olmasının gerekliliği de birçok araştırmacı tarafından vurgulanmaktadır (Aycan, 2000; Aycan, Kanungo, Mendonca, Yu, Deller, Stahl ve Kurshid, 2000; De Luque ve Sommer, 2000; Fletcher ve Perry, 2001; Levy ve Williams, 2004; Ramamoorthy ve Carroll, 1998; Sümer, 2000). Bu nedenle, çok kaynaklı değerlendirme gibi kökü batıya dayanan ve geleneksel olmayan değerlendirici kaynaklarının değerlendirme sürecine dahil edilmesini gerektiren bir yaklaşımın, kullanılacağı kültürel ortamla uyumluluğunun gösterilmesi bir zorunluluktur.

Bu çalışmanın amacı, geleneksel olmayan değerlendirici kaynaklarının, yönetici değerlendirmeleriyle birlikte performans değerlendirmesi sürecinde kullanılabilirliğini Türkiye bağlamında değerlendirmektir. Bu amaca yönelik olarak, hiyerarşik yapının belirgin olduğu bir kamu kuruluşunda kapsamlı bir performans değerlendirme sistemi çalışması bünyesinde, bir “çok kaynaklı değerlendirme” uygulaması yapılmıştır. Bu uygulamanın hedefi, geleneksel olmayan değerlendirici kaynakları tarafından yapılan değerlendirmelerin psikometrik özelliklerini ve geleneksel olmayan kaynakların kullanımına yönelik kullanıcı tutumlarını incelemektir. Ancak, bu uygulama pratik nedenlerden dolayı, yapılan çok kaynaklı değerlendirmelerin geribildirimini kapsamamıştır. Bu bağlamda, bu çalışmanın, çok kaynaklı geribildirim sistemlerinin tam bir testi olduğunu söylemek mümkün değildir. Çalışmada, farklı kaynakların yaptığı değerlendirmelerin ölçüm kalitesine ve kullanıcı tepkilerine odaklanılmıştır.

Yöntem

Örneklem

Çalışmanın örneklemini, güç mesafesinin

¹Bu makalede, çalışmanın amacıyla tutarlı olarak, amir değerlendirmeleri ile ilgili verinin tamamı hiçbir analizde kullanılmamış, sadece küçük bir alt seti kullanılmıştır.

oldukça belirgin olduğu, dikey örgütsel yapılanmanın gözlemlendiği geleneksel bir kamu kuruluşunda görevli personel arasından seçkisiz olarak seçilen yönetici pozisyonundaki kamu çalışanları oluşturmuştur. Performans değerlendirmesi veren ve alan tüm katılımcılar, araştırmanın gerçekleştirildiği kamu kurumunda alt, orta veya yüksek yöneticilik pozisyonlarında görev yapmaktadır. Diğer bir deyişle, aşağıda ast olarak tanımlanan katılımcıların kendileri de görece alt düzey yönetici pozisyonunda olan çalışanlardır.

Söz konusu kamu kurumunda gerçekleştirilen kapsamlı bir araştırmanın bir alt çalışması olan bu çalışmada, yapılan analizlerin amacına uygun olarak, örneklem sayıları (hem değerlendirme yapan hem de değerlendirilen kişi sayıları) değişmiştir. Bu makalede yapılan analizlerde her bir değerlendirici-değerlendirilen kombinasyonu için kullanılan temel örneklem sayıları şu şekildedir: 251 yönetici¹ (değerlendirilen 785 ast), 29 üst yönetici (değerlendirilen 101 ast), 302 çalışma arkadaşı (değerlendirilen 345 çalışan, yapılan değerlendirme sayısı 1638), 663 ast (değerlendirilen 118 yönetici). Ayrıca, 159 çalışan da öz değerlendirmesi yapmıştır. Yaşları 23 ile 55 arasında değişen katılımcılar arasında erkek katılımcıların oranı, kullanılan her bir alt örnekleme %90 veya üzerindedir. Değerlendirme veren ve alan tüm katılımcılar, araştırmanın gerçekleştirildiği kamu kurumunda alt, orta veya yüksek yöneticilik pozisyonlarında görev yapmaktadır.

Ölçüm Araçları

Performans değerlendirme formları. Bu çalışmada beş ayrı performans değerlendirme formu (yönetici formu, üst yönetici formu, öz değerlendirme formu, çalışma arkadaşı formu ve ast formu) kullanılmıştır. Yönetici, üst yönetici ve öz değerlendirme formları aynı 18 boyutu kapsamış, formların sadece yönergeleri farklılaşmıştır. Diğer bir deyişle, hem astlarını değerlendiren yöneticiler ve

üst yöneticiler hem de kendi performansını değerlendiren çalışanlar, değerlendirmelerini aynı 18 boyut üzerinde yapmıştır. Yönetici ve öz değerlendirmelerinde kullanılan boyutlar, hem görev performansına (“Yapılan işin kalitesi” gibi) hem de ortamsal performansa (“Uyumluluk” gibi) yönelik boyutlardır. Çalışma arkadaşlarının yaptıkları değerlendirmede kullanılan formda ağırlıklı olarak kişinin ortamsal performansını temsil eden (“İşbirliğine yatkınlık” gibi), ancak görev performansını da kapsayan altı boyut yer almıştır. Ast değerlendirmelerinde kullanılan formda ise yine ağırlıklı olarak ortamsal performansı (“Görev dağılımında hakkaniyet” gibi) ve sınırlı sayıda da görev performansını temsil eden 12 performans boyutu yer almıştır². Değerlendirmede kullanılan boyutlar çok aşamalı, iteratif bir yaklaşımla belirlenmiştir. Öncelikle kurum içinde halihazırda kullanılan değerlendirme araçları ve var olan iş tanımları proje ekibi tarafından incelenmiştir. Daha sonra, var olan ve önerilen sistemlerin değerlendirilmesine yönelik potansiyel kullanıcılarla 18 ayrı odak grubu yapılmıştır. Kapsamlı bir anket çalışması (N > 3000) aracılığı ile de içerilecek boyutlara yönelik kullanıcı tepkileri değerlendirilmiştir. Tüm bu aşamalardan sonra tespit edilen boyutlar, kurum içinden ve proje grubundan uzmanlar tarafından değerlendirilmiş, birleştirmeler ve elemeler sonunda boyutlar son hallerini almıştır. Boyut bazında yapılan tüm bu değerlendirmeler 5-basamaklı Likert tipi (1 = Yetersiz Performans; 5 = Üstün Performans) bir ölçek üzerinde yapılmıştır. Farklı kaynakların kullandıkları bu formların psikometrik özellikleri bulgular bölümünde ele alınmaktadır.

Kullanıcı tepkisi anketleri. Yönetici, çalışma arkadaşı, ast veya öz olarak performans değerlendirmesi yapan katılımcıların, çok kaynaklı değerlendirmeye yönelik tutumları da değerlendirilmiştir. Çok kaynaklı performans değerlendirmesine yönelik kullanıcı tepkilerini ölçmeyi hedefleyen

anketlerde katılımcılardan, temsil ettikleri değerlendirici kaynağının çok kaynaklı bir sistem içinde yer alıp almaması gerektiğine yönelik maddeleri (örn., “Ast değerlendirmelerinin amir-çalışan ilişkisine zarar vereceğini düşünüyorum” ve “Çalışma arkadaşlarım için yaptığım değerlendirmelerin, kimliğimin gizli kalması koşuluyla, kendilerine iletmesini isterim”) 5-basamaklı Likert tipi bir ölçek üzerinde (1 = Hiç Katılmıyorum; 5 = Tamamen Katılıyorum) değerlendirmeleri istenmiştir. Bu makalede, sadece çalışma arkadaşı, ast ve/ya öz olarak değerlendirme yapan kişilerin doldurduğu kullanıcı tepkisi anketlerinde, söz konusu kaynağa yönelik tutum maddelerine verilen cevaplar analiz edilmiştir. Çalışma arkadaşı ve ast değerlendirmeleri için 9’ar madde, öz değerlendirmeleri için ise 4 madde analize dahil edilmiştir. Her bir grup madde öncelikle faktör analizine tabi tutulmuştur. Bulgular bölümünde açıklandığı gibi, bazı maddeler faktörler altında gruplanırken, gruplanmayan maddeler tek tek ele alınmıştır. Tablo 5-7’de bu anketlerde yer alan maddeler (ya da madde grupları) yer almaktadır.

İşlem

Çalışmanın yapıldığı kamu kuruluşunda, öncelikle uygulamaların yapılacağı bölgeler ve birimler belirlenmiş, daha sonra ise her bir bölgeden gelen temsilcilere kısa bir eğitim verilmiştir. Bu eğitimin amacı kullanılacak çok kaynaklı değerlendirme sistemini tanıtmak ve uygulama sırasında ortaya çıkabilecek problemlere karşı bölge temsilcilerini hazırlamaktır. Öz, yönetici, üst yönetici, çalışma arkadaşı ve/ya ast olarak değerlendirme yapacak kişilere, uygulama öncesinde, değerlendirmede dikkat etmeleri gereken noktalara yönelik yazılı yönergeler verilmiş ve yapacakları değerlendirmelerin sadece araştırma amaçlı kullanılacağı belirtilmiştir. Bazı katılımcılar, sadece bir değerlendirici kaynağını temsil ederken, bazıları ise birden fazla kayna-

²Kullanılan formların geçerliliğini ve gizliliğini korumak amacıyla formlarda yer alan boyutlar bu makalede paylaşılmamaktadır. Bu çalışmanın hedefinin, çok kaynaklı değerlendirme yaklaşımının psikometrik açıdan ve kullanıcı tepkileri temelinde değerlendirilmesi olduğu düşünüldüğünde, boyutların tek tek paylaşılmasının, bir engel oluşturmayacağı düşünülmüştür.

ğı temsil ederek (örn., hem ast, hem de çalışma arkadaşı olarak), değerlendirmeler yapmıştır. Ayrıca, yönetici, üst yönetici ve çalışma arkadaşı olarak değerlendirme yapan katılımcılar çoğunlukla birden fazla kişiyi değerlendirmişlerdir. Yönetici, üst yönetici, çalışma arkadaşı veya ast olarak değerlendirme yapacak kişilere, değerlendirecekleri kişinin/kişilerin isimleri önceden sağlanmıştır. Yapılan performans değerlendirmelerinin ardından, değerlendiricilerden yaptıkları değerlendirmelere ve temsil ettikleri değerlendirici kaynağının değerlendirme sisteminde yer alıp almamasına yönelik tutumlarını ölçen, kullanıcı tepkisi anketi uygulanmıştır. Yapılan performans değerlendirmeleri ve doldurulan kullanıcı tepkileri anketleri, bölge temsilcileri aracılığı ile araştırmacılara ulaştırılmıştır.

Bulgular

Çalışmanın amacı ile tutarlı olarak, veriler üzerinde iki grup analiz yapılmıştır. İlk olarak, farklı kaynakların yaptıkları değerlendirmeler psikometrik açıdan karşılaştırılmıştır. İkinci olarak ise, geleneksel olmayan kaynakların, performans yönetimi sisteminde yer almasına yönelik kullanıcı tepkileri betimsel düzeyde irdelenmiştir.

Farklı Kaynaklar Tarafından Yapılan Performans Değerlendirmelerinin Psikometrik Açısından Değerlendirilmesi

Bu amaçla yapılan analizleri üç başlık altında toplamak mümkündür. İlk olarak, değerlendirmelerde kaynak içi ve kaynaklar arası korelasyonlar incelenmiştir. İkinci olarak, değerlendirmelerde kaynak etkisini incelemek amacıyla, aynı kişiyi değerlendiren farklı değerlendirici gruplarının yaptıkları değerlendirmelerin karşılaştırıldığı varyans analizleri sunulmuştur. Bu analizler ve izleyen ikili karşılaştırmalar, değerlendirici kaynakları temelinde cömertlik etkisinin (olumlu uçta değerlendirme yapma eğilimi) incelenmesine izin vermiştir. Son olarak ise, hale etkisini (değerlendirme yapılan boyutları birbirlerinden ayıramama eğilimi) incelemek amacı ile farklı değerlendirici kaynaklarının yaptıkları değerlendirmeler için boyutlar arası ko-

relasyonlar ve kullanılan formların iç tutarlık katsayıları incelenmiştir.

Kaynak içi ve kaynaklar arası korelasyonlar. Kaynak içi (intraclass – ICC) korelasyonlar, aynı kişiyi değerlendiren tek bir değerlendirici kaynağının (örn., astların ya da denklemin) kendi içindeki hemfikir/tutarlı olma derecesini yansıtan bir indekstir (Garson, 2005). Bu çalışmada, kaynak içi korelasyonların hesaplanmasındaki amaç, iki yönlü varyans analizini kullanarak aynı kişiyi değerlendiren değerlendiriciler arası hemfikir olma derecesini tespit edebilmektir. Yapılan analizlerde, Shrout ve Fleiss (1979) tarafından önerilen üç ICC modelinden üçüncü model olan iki yönlü karmaşık etki modeli (“two-way mixed design”) kullanılmıştır. Bu analizlerde, aynı kişiyi aynı boyutlarda değerlendiren değerlendiriciler, “sabit” (“fixed”) etkiye, yapılan boyut değerlendirmeleri de “seçkisiz” (“random”) etkiye karşılık gelmektedir (ayrıntılı bilgi için bkz. Yaffee, 1998). Ast ve çalışma arkadaşları için ortalama kaynak içi korelasyonun (değerlendiriciler arası ortalama tutarlılık katsayısının) hesaplanmasında 2 aşamalı bir yaklaşım kullanılmıştır. Öncelikle, değerlendiriciler arası tutarlılık katsayısı, en az 5 astı ya da çalışma arkadaşı tarafından değerlendirilen her bir kişi için ayrı ayrı hesaplanmıştır. Daha sonra ise söz konusu değerlendirme kaynağına genellenebilecek değerlendiriciler arası iç tutarlık katsayısını belirleyebilmek amacı ile o kaynak tarafından değerlendirilen her bir kişi için hesaplanmış olan ICC değerlerinin ortalama, ortanca ve standart sapmaları hesaplanmıştır.

Toplam 345 çalışan için 1638 çalışma arkadaşı değerlendirmesi yapılmıştır. Değerlendirme yapan çalışma arkadaşı sayısı 302’dir. Aynı kişi için değerlendirme yapan çalışma arkadaşı sayısı 1-13 arasında değişmiştir. Kaynak içi ortalama hemfikir olma derecesini tespit edebilmek amacıyla, yukarıda da belirtildiği gibi en az 5 çalışma arkadaşı tarafından performansı değerlendirilen her bir kişi için, kaynak içi korelasyonlar hesaplanmıştır. Bu korelasyonların ortalaması .28 (standart sapma = .44, ortanca korelasyon =.38) olarak bulunmuştur. Bu

ortalama, 71 ayrı çalışma arkadaşı grubunun yaptığı değerlendirmeleri kapsamaktadır (bir değerlendirici, birden fazla çalışma arkadaşı değerlendirme grubunda yer alabilmiştir). Standart sapmanın görece büyüklüğünden anlaşılacağı gibi, bazı kişiler için çalışma arkadaşlarının yaptığı değerlendirmelerin arasındaki korelasyon çok yüksek ve pozitifken, bir çok kişi için korelasyon 0'a yakın ya da negatif olarak gözlenmiştir. Aynı işlem ast değerlendirmeleri için tekrarlanmıştır. Çalışmada, 663 ast, 118 yöneticisini değerlendirmiştir. Aynı yönetici için değerlendirme sağlayan ast sayısı 1-37 arasında değişmiştir. Ast değerlendirmeleri için kaynak içi ortalama korelasyon katsayısı .49 (standart sapma = .37, ortanca korelasyon = .58) olarak bulunmuştur. Bu ortalama, 42 ast grubunun yaptığı yönetici değerlendirmelerini kapsamıştır.

Aynı kişiyi değerlendiren farklı değerlendirici grupları arasında hemfikir olma derecesini değerlendirmek amacıyla kaynaklar arası korelasyonlar incelenmiştir. Tablo 1'de aynı çalışanlar için farklı kaynakların sağladığı değerlendirmelerin (ortalama performans puanı üzerinden) korelasyonları verilmiştir. Tablo 1'de görüldüğü gibi, yönetici-üst yönetici (.58), ast-yönetici (.26), ast-üst yönetici

(.33), çalışma arkadaşı-öz (.30), çalışma arkadaşı-yönetici (.31) ve çalışma arkadaşı-üst yönetici (.37) arasındaki korelasyonlar anlamlıdır. Yönetici-üst yönetici değerlendirmeleri arasındaki korelasyon en yüksek iken, en düşük korelasyonlar öz-üst yönetici ve çalışma arkadaşı-ast değerlendirmeleri arasındaki korelasyonlardır.

Kaynaklar arası karşılaştırmalar. Tablo 2'de aynı kişi için farklı kaynaklar tarafından yapılan değerlendirmelerin karşılaştırıldığı, tekrar ölçümlü varyans analizlerinin sonuçları ve değerlendirme ortalamaları yer almaktadır. Bu analizler kişilere verilen ortalama puanlar (tüm performans boyutlarında verilen puanların toplanıp boyut sayısına bölünmesiyle elde edilen puan; minimum = 1, maksimum = 5) üzerinden yapılmıştır. Analizlerde, birden fazla değerlendiricinin bulunduğu durumlarda (çalışma arkadaşları ve ast değerlendirmelerinde) kaynak ortalaması alınmıştır. İlk olarak, 5 kaynağın (öz, yönetici, üst yönetici, astlar ve çalışma arkadaşlarının) aynı kişi için verdiği değerlendirmelerin karşılaştırıldığı analiz sunulmuştur. Tüm kaynaklar tarafından değerlendirilen kişi sayısı (N = 31) görece küçüktür. İkinci olarak, yalnızca yönetici, çalışma arkadaşları ve astlar tarafından değerlendirilen

Tablo 1

Farklı Kaynaklar (Öz, Yönetici, Üst Yönetici, Ast ve Çalışma Arkadaşları) Tarafından Yapılan Değerlendirmeler Arası Korelasyonlar

| | Öz | Yönetici | Üst Yönet. | Ast | Çal. Ark. |
|--------------|----------------|----------------|----------------|----------------|----------------|
| Öz | 1.000 (166) | | | | |
| Yönetici | .17 (109) | 1.000 (339) | | | |
| Üst Yönetici | -.05 (51) | .58** (64) | 1.000 (100) | | |
| Ast | .17 (56) | .26* (74) | .33* (41) | 1.000 (118) | |
| Çal. Ark. | .30** (113) | .31** (256) | .37* (45) | .05 (59) | 1.000 (345) |

**p < .01 , *p < .05. Parantez içindeki değerler, söz konusu korelasyonun alındığı örneklemin büyüklüğüne karşılık gelmektedir.

Tablo 2

Farklı Kaynaklar (Öz, Yönetici, Üst Yönetici, Ast, ve Çalışma Arkadaşı) Tarafından Yapılan Değerlendirmelerin Tekrar Ölçümlü Varyans Analizleri Sonuçları ve Her bir Kaynak İçin Ortalamalar ve Standart Sapmalar

| Değerlendirme Kaynakları | Hata sd | Grup sd | F | p | η^2 | Öz Ort. (SS) | Yönetici Ort. (SS) | Üst Y. Ort. (SS) | Ast Ort. (SS) | Çal. A. Ort. (SS) |
|----------------------------------------|---------|---------|-------|------|----------|---------------|--------------------|------------------|---------------|-------------------|
| Öz, Yönetici, Üst Yön., Ast, Çal. Ark. | 120 | 4 | 13.61 | .001 | .31 | 4.75 (.30) | 4.00 (.71) | 4.22 (.63) | 4.44 (.60) | 4.67 (.33) |
| Yönetici, Ast, Çal. Ark. | 90 | 2 | 8.99 | .001 | .17 | - | 4.14 (.72) | - | 4.38 (.58) | 4.60 (.42) |
| Yönetici, Üst Yönetici | 63 | 1 | 5.73 | .020 | .08 | - | 4.29 (.69) | 4.47 (.57) | - | - |

Karşılaştırmalar toplam ortalama puanlar (5-basamaklı ölçek) üzerinde yapılmıştır.

kişiler (N = 46) için yapılan değerlendirmelerin karşılaştırıldığı varyans analizinin sonuçları sunulmuştur. Son olarak ise, aynı çalışanlar (N = 64) için yapılan yönetici ve üst yönetici değerlendirmeleri karşılaştırılmıştır.

Öncelikle, farklı kaynakların değerlendirme ortalamaları incelendiğinde, tüm kaynakların cömertçe, ölçeğin olumlu ucunu kullanarak değerlendirme yaptığı görülmektedir. Beş kaynağın karşılaştırıldığı varyans analizi sonuçlarına göre, aynı kişi için farklı kaynakların yaptıkları değerlendirmeler ara-

sındaki fark istatistiksel olarak anlamlıdır ve güçlüdür ($\eta^2 = .31$). En yüksek değerlendirme öz değerlendirmesi iken, en düşük değerlendirme yönetici değerlendirmesidir. Tablo 3'te anlamlı bulunan bu farklılıkları irdeleyen ikili karşılaştırma sonuçları yer almaktadır. Karşılaştırma sonuçlarına göre, yönetici değerlendirmeleri, ast, çalışma arkadaşları ve öz değerlendirmelerinden anlamlı derecede daha düşüktür. Üst yönetici değerlendirmeleri ise öz ve çalışma arkadaşı değerlendirmelerinden anlamlı derecede daha düşüktür.

Tablo 3

Farklı Kaynaklar (Yönetici, Öz, Çalışma Arkadaşları ve Astar) Tarafından Yapılan Değerlendirmeler Üzerinde Yapılan Varyans Analizleri Sonucunda Tespit Edilen Anlamlı Farklılıklar

| Değerlendirme Kaynağı | Öz | Yönetici | Üst Yönetici | Ast | Çalışma Arkadaşı |
|------------------------------------|------------------------|--------------------------------------|----------------------------------|----------|--------------------------------------|
| Öz, Yön., Üst Yön., Ast, Çal. Ark. | Öz>Üst Yön. Öz>Yön. | Yön.<Ast Yön.<Çal. Ark Yön.<Öz | ÜstYön.<Çal. Ark. Üst Yön.<Öz | Ast>Yön. | Çal. Ark.>Üst Yön. Çal. Ark.>Yön. |
| Yön., Ast, Çal. Ark. | | Yönetici<Çal. Ark. | | | Çal. Ark.>Yön. |

Tüm farklılıklar p < .01'de anlamlıdır.

İkinci varyans analizinde (Tablo 2) aynı kişi için yapılan yönetici, çalışma arkadaşları ve ast değerlendirmeleri karşılaştırılmıştır. Analiz sonuçlarına göre, bu üç kaynak arasında anlamlı bir farklılık ($\eta^2 = .17$) gözlenmiştir. Bu veri setinde, en yüksek değerlendirme çalışma arkadaşları tarafından yapılırken, en düşük değerlendirme yöneticiler tarafından yapılmıştır. Tablo 3'te sunulan ikili karşılaştırmalar sonunda yönetici değerlendirmelerinin çalışma arkadaşı değerlendirmelerinden anlamlı olarak daha düşük olduğu bulunmuştur.

Son analizde ise, aynı çalışan için yapılan yönetici ve üst yönetici değerlendirmeleri karşılaştırılmıştır. Analiz sonuçlarına göre, üst yönetici değerlendirmeleri, yönetici değerlendirmelerinden daha yüksektir ($\eta^2 = .08$).

Boyutlar arası korelasyonlar ve formlar için iç tutarlık katsayıları. Yukarıda da belirtildiği üzere, farklı kaynaklar, farklı formlar kullanarak değerlendirme yapmıştır. Bunun nedeni, aynı kişiyi değerlendirme pozisyonunda olabilecek farklı kaynakların, performansın farklı yönlerini daha etkili (ya da etkisiz) değerlendirecekleri beklentisidir. Bu nedenle, farklı kaynaklar için hazırlanan değerlendirme formlarında ortak bazı boyutlar yer alsa da, genel olarak kapsanan boyutlar, boyutların içeriği ve/ya sayısı farklılaşmıştır. Aşağıda sunulan analizlerin amacı, farklı kaynakların kullandığı formların iç tutarlık katsayılarını ve içerilen boyutlar arasındaki korelasyon katsayılarını incelemektir.

Tablo 4

Farklı Kaynakların Yaptıkları Değerlendirmelerin İç Tutarlık Katsayıları ve Kullanılan Formlarda Boyutlar Arası Korelasyonlar

| Değerlendirme Kaynağı | Boyut Sayısı | N | İç-Tutarlık Katsayısı | Boyutlar Arası Ortalama Korelasyon |
|-----------------------|--------------|------|-----------------------|------------------------------------|
| Yönetici | 18 | 57 | .97 | .67 |
| Üst Yönetici | 18 | 99 | .98 | .78 |
| Çalışma Arkadaşları | 7 | 1597 | .96 | .77 |
| Astlar | 12 | 646 | .97 | .75 |
| Öz | 18 | 161 | .95 | .57 |

N'ler her bir değerlendirici grubu tarafından yapılan değerlendirme sayısına karşılık gelmektedir.

Tablo 4'te görüleceği gibi, tüm formlar için iç tutarlık katsayısı oldukça yüksektir. Bu bulgu, içerilen boyutların birbiri ile örtüştüğü anlamına gelebileceği gibi, değerlendiricilerin boyutlar arası ayırım yapma konusunda istekli ya da yetkin olmamalarından da kaynaklanabilir. Diğer bir deyişle, yüksek iç tutarlık katsayısı, hale etkisinin (kavramsal olarak ayrı olan performans boyutlarını birbirinden ayırt edememe) bir göstergesi olabilir. Hale etkisinin varlığını incelemede kullanılacak bir diğer önemli kriter, boyutlar arası korelasyonlardır. Görüleceği gibi, boyutlar arası korelasyon, en düşük öz değerlendirmeleri için, daha sonra yönetici değerlendirmeleri için tespit edilmiştir. Üst yöneticiler, çalışma arkadaşları ve astların yaptıkları değerlendirmelerde boyutlar arası korelasyonlar görece yüksektir.

Geleneksel Olmayan Değerlendirici Kaynaklarına Yönelik Tutumların Değerlendirilmesi

Çok kaynaklı değerlendirme uygulamasına katılan katılımcılardan, uygulamanın hemen ardından, yaptıkları değerlendirmelerin farklı yönlerine yönelik (yönetici, çalışma arkadaşı ve/ya öz değerlendirmelerine) görüşlerini 5 basamaklı Likert tipi ölçekler üzerinde belirtmeleri istenmiştir. Kullanıcı tepkisi anketinde yer alan ve geleneksel olmayan değerlendirme kaynaklarına yönelik olan maddeler farklı bir seri temel bileşenler analizine tabi tutulmuş³, bu analizler sonunda gruplaştıkları görülen bazı maddeler için iç tutarlık katsayıları hesaplan-

Tablo 5**Çalışma Arkadaşları Değerlendirmesine Yönelik Tutumlara İlişkin Betimleyici İstatistikler**

| İfade | N | Ort. | Std. Sp. |
|--------------------------------------------------------------------------------------------------------------------------------|-----|------|----------|
| 1 Çalışma arkadaşlarımı değerlendirmek zor geldi. (Madde sayısı = 3, Alfa = .80) | 293 | 2.24 | 1.32 |
| 2 Çalışma arkadaşı değerlendirmeleri, değerlendirme sisteminde yer almalıdır. (Madde sayısı = 3, Alfa = .85) | 292 | 3.67 | 1.33 |
| 3 Çalışma arkadaşı değerlendirmeleri sadece geri besleme amacı ile kullanılmalıdır. | 293 | 3.44 | 1.62 |
| 4 Çalışma arkadaşı değerlendirmelerinin iş ilişkilerine zarar vereceğini düşünüyorum. | 292 | 2.77 | 1.61 |
| 5 Çalışma arkadaşlarım için yaptığım değerlendirmelerin, kimliğimin gizli kalması koşuluyla, kendilerine iletilmesini isterim. | 292 | 3.87 | 1.53 |

Değerlendirme 5-basamaklı bir ölçek üzerinde yapılmıştır (1 = Hiç Katılmıyorum; 5 = Tamamen Katılıyorum).

mıştır. Herhangi bir faktör altında yer almayan maddeler için betimleyici analizler ise madde bazında yapılmıştır. Tablo 5'te çalışma arkadaşı değerlendirmesini yapan 293 kişinin, bu değerlendirme kaynağının performans yönetimi sisteminde yer almasına yönelik görüşleri ile ilgili temel betimleyici istatistikler sunulmuştur. Görüleceği gibi, çalışma arkadaşlarını değerlendirmek katılımcılara çok zor gelmediği gibi (Ort = 2.24), çok kolay da

gelmemiştir. Çalışma arkadaşlarının değerlendirmelerinin, değerlendirme sisteminde yer almasına (Ort = 3.67) ve de sadece geribildirim amaçlı kullanılmasına (Ort = 3.44) yönelik olarak ise görece olumlu tutumların olduğu gözlenmiştir. İş arkadaşlarının birbirlerini değerlendirmesinin iş ilişkilerine zarar vereceği görüşüne (Ort = 2.77), ortanın hemen altında bir düzeyde bir katılım olduğu görülmüştür. Çalışma arkadaşlarının yaptıkları değeren-

Tablo 6**Ast Değerlendirmesine Yönelik Tutumlara İlişkin Betimleyici İstatistikler**

| İfade | N | Ort. | Std. Sp. |
|----------------------------------------------------------------------------------------------------------------|-----|------|----------|
| 1 Amirimi değerlendirmek zor geldi. (Madde sayısı = 2, Alfa = .70) | 660 | 2.09 | 1.26 |
| 2 Ast değerlendirmeleri sistem içinde yer almalıdır. (Madde sayısı = 3, Alfa = .75) | 656 | 4.11 | 1.01 |
| 3 Astlar tarafından yapılan değerlendirmeler sadece geri besleme amacı ile kullanılmalıdır. | 660 | 2.47 | 1.64 |
| 4 Ast değerlendirmelerinin amir-çalışan ilişkisine zarar vereceğini düşünüyorum. | 660 | 2.29 | 1.53 |
| 5 Amirim için yaptığım değerlendirmelerin, kimliğimin gizli kalması koşuluyla, kendisine iletilmesini isterim. | 658 | 3.91 | 1.47 |

Değerlendirme 5-basamaklı bir ölçek üzerinde yapılmıştır (1 = Hiç Katılmıyorum; 5 = Tamamen Katılıyorum).

³Çalışma arkadaşlarının değerlendirmelerine yönelik tutumlarla ilgili maddeler üzerinde yapılan varyans analizi sonunda, varyansın %60'ını açıklayan "değerlendirmelerin yararı" ve "değerlendirmelerin zorluğu" olarak adlandırılabilir iki faktörün olduğu, diğer maddelerin ise herhangi bir faktör altında yer almadığı görülmüştür. Benzer şekilde adlandırılan ve varyansın %53.4'ünü açıklayan iki faktör de ast değerlendirmelerine yönelik tutumlar için tespit edilmiştir.

Tablo 7**Öz Değerlendirmesine Yönelik Tutumlara İlişkin Betimleyici İstatistikler**

| İfade | N | Ort. | Std. Sp. |
|---------------------------------------------------------------------------------------------------------------|-----|------|----------|
| 1. Yaptığım bu değerlendirmenin, zayıf ve güçlü yönlerimi iyi yansıttığını düşünüyorum. | 158 | 4.16 | 1.25 |
| 2. Bir kişi kendisini tarafsız değerlendirebilir. | 159 | 3.06 | 1.49 |
| 3. Kendimi değerlendirmek bana zor geldi. | 159 | 2.86 | 1.61 |
| 4. Kendim için yaptığım bu değerlendirmenin amirim beni değerlendirmesinden daha olumlu olduğunu düşünüyorum. | 158 | 3.37 | 1.44 |

Değerlendirme 5-basamaklı bir ölçek üzerinde yapılmıştır (1 = Hiç Katılmıyorum; 5 = Tamamen Katılıyorum).

dirmelerin, gizliliğin sağlanması koşuluyla, değerlendirilen kişilere geribildirim yapılmasına yönelik, ortanın üzerinde (Ort = 3.87) olumlu bir tutum gözlenmiştir.

Tablo 6’da yöneticisini değerlendiren katılımcıların (N = 660), bu değerlendirmelere yönelik tutum sorularına verdikleri cevaplarla ilgili betimleyici istatistikler yer almaktadır. Öncelikle, yöneticilerini değerlendirmenin, katılımcılara çok zor gelmediği (Ort = 2.09) görülmektedir. Ast değerlendirmelerinin, performans yönetimi sistemi içinde yer almasına yönelik oldukça olumlu (Ort = 4.11) bir tutum gözlenmiştir. Katılımcıların, ast değerlendirmelerinin sadece geribildirim amaçlı kullanımı fikrine çok fazla katılmadıkları (Ort = 2.47) tespit edilmiştir. Benzer şekilde, katılımcılar, astların yaptıkları değerlendirmelerin, yönetici-ast ilişkisine zarar vereceği fikrine de çok katılmamışlardır (Ort = 2.29). Son olarak, değerlendirme yapan astın kimliğinin korunması koşuluyla, yapılan değerlendirmelerin, değerlendirilen yöneticiye iletilmesi konusunda oldukça (Ort = 3.91) olumlu bir tutum gözlenmiştir.

Tablo 7’de öz değerlendirmelerine katılan katılımcıların (N = 159), bu değerlendirmelere yönelik tutumlarıyla ilgili betimleyici istatistikler yer almaktadır. Görüleceği gibi, öz değerlendirmesi yapan katılımcılar, bu değerlendirmelerde performanslarını oldukça gerçekçi bir şekilde (Ort = 4.16) yansıttıklarını düşünmektedir. Bununla bera-

ber katılımcılar, genel olarak kişilerin kendilerini yansız değerlendirebileceği görüşüne (Ort = 3.06) orta düzeyde katılmışlardır. Katılımcıların, öz değerlendirmelerinin zor olduğu görüşüne ise ortaya yakın (Ort = 2.86) düzeyde katıldıkları görülmüştür. Son olarak, öz değerlendirmelerinin, amirlerin yaptığı değerlendirmelerden daha “cömertçe” olduğu görüşüne de ortanın üzerinde katıldıkları bulunmuştur (Ort = 3.37).

Tartışma

Bu çalışmanın amacı bir kamu kuruluşunda gerçekleştirilen çok kaynaklı değerlendirme uygulamasını ölçüm kalitesi ve kullanıcı tepkileri açısından incelemektir. Yapılan analizler sonunda, astlar için kaynak içi korelasyon katsayılarının, Conway ve Huffcutt’ın (1997) meta analiz çalışmasında rapor ettikleri katsayılardan yüksek olduğu bulunmuştur. Ayrıca, yine yazından farklı olarak, ast değerlendirmeleri için tespit edilen ortalama kaynak içi korelasyon katsayısının, çalışma arkadaşları için tespit edilen ortalama katsayıdan daha yüksek olduğu da gözlenmiştir. Bu bulgu, özellikle astların yaptığı değerlendirmelerin geçerliğinin bir işareti olarak yorumlanabilir. Çalışma arkadaşı değerlendirmelerinde gözlenen görece düşük kaynak içi korelasyonlar ve yüksek standart sapma ise, ast değerlendirmelerine kıyasla çalışma arkadaşı değerlendirmelerinin geçerliğinin görece düşük olduğunu düşündürmektedir.

Bu çalışmada kaynaklar arasında gözlenen korelasyonların örüntüsü, yazında rapor edilen benzer değerlendirici kaynakları arasındaki korelasyonların örüntüsüne paralellik göstermektedir. Örneğin, yukarıda da bahsedildiği gibi, Harris ve Schaubroeck'ın (1988) farklı değerlendirme kaynaklarına yönelik meta-analiz çalışmasında; çalışma arkadaşı, yönetici ve öz değerlendirmeleri karşılaştırılmıştır. Harris ve Schaubroeck'ın çalışmasında kullanılan bu üç kaynağın birbirleri arasındaki korelasyonların sıralaması, bu çalışmada elde edilen korelasyonların sıralamasıyla aynıdır. Bununla birlikte, Türk örneklemeden elde edilen korelasyonlar görece daha düşüktür. Gözlenen korelasyonların düşük olmasının önemli bir nedeni, aşağıda açıklandığı gibi, değerlendirmelerin dar bir aralık içinde (cömertçe) yapılmış olması olabilir. Değerlendirme aralığının dar olmasının gözlenen korelasyonları küçülteceği bilinmektedir (bkz., Tsui ve Ohlott, 1988).

Gözlenen kaynaklar arası korelasyonların örüntüsü ve büyüklükleri, bir istisna (ast-çalışma arkadaşı çifti) dışında Conway ve Huffcutt (1997) ve Atkins ve Wood (2002) tarafından rapor edilen korelasyonlarla da büyük oranda benzeşmektedir. Yaptığımız çalışmada, ast-çalışma arkadaşı değerlendirmeleri arasındaki korelasyon (.05), Conway ve Huffcutt'un meta-analizlerinde rapor ettiği (.22) aynı değerlendirici grupları arasındaki korelasyondan oldukça düşüktür. Bu bulgu, yapılan çalışmada ast ve çalışma arkadaşı olarak değerlendirme yapan kişilerin, performans değerlendirmeleri yaparken muhtemelen farklı perspektiflerden baktıkları şeklinde yorumlanabilir. Ayrıca, özellikle çalışma arkadaşı değerlendirmelerinin cömertlik etkisinin en fazla gözlendiği değerlendirmeler arasında olduğu, dolayısıyla da dar bir aralıkta olduğu düşünüldüğünde, gözlenen düşük korelasyon çok şaşırtıcı değildir. Alternatif olarak, gözlenen bu düşük korelasyonun, ast ve çalışma arkadaşı değerlendirmelerinde tamamen aynı boyutların kullanılmamasından kaynaklandığı ileri sürülebilir. Bununla beraber, bu iki kaynak tarafından kullanılan boyutların

farklı olmasının, gözlenen düşük korelasyonu açıklamada tek başına yeterli olmayacağı da açıktır. Ast ve çalışma arkadaşlarının yaptığı değerlendirmelerde kullanılan boyutların sayısı farklı olsa da, bu boyutların temsil ettiği performans alanları (ortamsal ve görev) aynı kalmıştır. Ayrıca, kaynaklar arası korelasyonların, tüm boyut değerlendirmelerinin ortalamasını yansıtan ortalama performans puanı kullanılarak hesaplandığı düşünüldüğünde de, boyut bazındaki bazı farklılaşmaların gözlenen düşük korelasyonlardan tamamen sorumlu olmayacağı görülmektedir.

Farklı değerlendirme kaynaklarının aynı kişiler için yaptıkları toplam performans puanı değerlendirmeleri üzerinde yapılan analizler sonunda, tüm kaynakların cömertçe, yani ölçeğin olumlu ucunu kullanarak değerlendirme yaptığı tespit edilmiştir. Kaynak karşılaştırmaları sonunda, en cömertçe yapılan değerlendirmelerin öz ve çalışma arkadaşı değerlendirmeleri olduğu, görece en düşük değerlendirmelerin ise yöneticiler tarafından verildiği gözlenmiştir. İki yönetici grubunun yaptığı değerlendirmeler karşılaştırıldığında ise, üst yöneticilerin, ilk amirlerden daha yüksek değerlendirme yapma eğiliminde olduğu görülmüştür. Gözlenen bu farklılıklara rağmen, vurgulanması gereken önemli bir nokta, tüm değerlendirici gruplarının cömertlik etkisine maruz olduklarıdır. Çalışma, araştırma amaçlı olmasına ve katılımcılar bu konuda ayrıntılı bir şekilde bilgilendirilmelerine rağmen, cömertlik etkisi yaygın bir şekilde tüm değerlendirmelerde gözlenmiştir. Gözlenen cömertlik etkisini kısmen de olsa, çalışılan kamu kurumunun değerlendirme kültürü üzerinde muhtemelen kuvvetli bir etkiye sahip olan ulusal kültürümüzle açıklamak mümkün olabilir. Fletcher ve Perry (2001), bizimki gibi toplulukçu kültürlerde (bkz., Hofstede, 1980, 1991) olumsuz geribildirimden kaçınmanın daha yüksek olduğunu ve değerlendirmelerin daha cömertçe yapıldığını gösteren görgül çalışmalara işaret etmektedir. Kültürün toplulukçu özelliğinin giderek azaldığını vurgulayan araştırmacılar olsa da (örn., Göregenli 1995, 1997), yaşadığımız

kültürde geleneksel ve toplulukçu öğelerin hala belirgin olduğunu gösteren yakın zamanda yapılmış çalışmalar da bulunmaktadır (örn., Aycan ve ark., 2000; Schwartz, akt., Triandis ve Bhawuk, 1997). Bu eğilimin, bu çalışmada yer alan kamu kuruluşu gibi kurumlarda, özel sektör kurumlarından daha belirgin olması da muhtemeldir. Bu değerlendirmeyle tutarlı olarak, Olmez, Sümer ve Soysal (2004), Türkiye’de kamu kurumlarının, özel sektör kurumlarına ve Türkiye’de faaliyet gösteren çok uluslu kurumlara oranla, daha geleneksel, bireyciliği karakterize eden rasyonellikten görece uzak bir çalışma anlayışını benimsediğini bulmuşlardır.

Yönetici ve çalışma arkadaşı değerlendirmeleriyle yüksek korelasyon göstermelerine karşın, üst yönetici değerlendirmeleri ilk amir değerlendirmelerinden daha yüksek olma eğilimindedir. Bu durumda, ilk amire kıyasla çalışan hakkında görece daha az bilgi sahibi olması beklenen üst yöneticilerin, değerlendirme sistemlerinde yer almamalarının büyük bir eksiklik olmayacağı söylenebilir.

Hale etkisine yönelik analiz sonuçlarına göre, öz değerlendirmelerinin hale etkisine daha az maruz olduğu söylenebilir. Üst yönetici, çalışma arkadaşı ve ast değerlendirmelerinin ise hale etkisine özellikle maruz oldukları gözlenmiştir. Kişilerin, kendi performanslarının farklı yönlerinin, performansı dışarıdan gözleyen ve değerlendiren kişilerden daha fazla farkında oldukları, diğer bir deyişle öz değerlendirmelerinin hale etkisine daha az açık olduğu yazında da (örn., Frone, Adams, Rice ve Instone-Noonan, 1986) rapor edilmektedir.

Geleneksel olmayan kaynaklara yönelik tutumlar incelendiğinde çalışma arkadaşlarının değerlendirme sisteminde yer almasına yönelik genel olarak temkinli bir tutumun olduğu söylenebilir. Çalışma arkadaşlarını değerlendirmeyi ne çok kolay ne de çok zor bulan katılımcılar, bu değerlendirmelerin sadece geribildirim amaçlı kullanılmasına (idari kararlara esas teşkil edecek şekilde değil) yönelik görece olumlu bir yaklaşım sergilemişlerdir. Ayrıca çalışma arkadaşı değerlendirmelerinin iş ilişkile-

rine zarar verebileceği görüşüne katılım da orta düzeye yakındır.

Ast değerlendirmesi yapan katılımcıların ast değerlendirmesine yönelik tepkileri incelendiğinde, öncelikle bu değerlendirmelere yönelik oldukça olumlu bir atmosferin olduğu ve amirlerini değerlendirmenin katılımcılara çok zor gelmediği görülmüştür. Ast değerlendirmelerinin kişilerin kimliği gizli kalmak koşuluyla değerlendirilen yöneticilere iletilmesine sıcak bakmakla birlikte, katılımcılar astların yaptıkları değerlendirmelerin sadece geribildirim amacıyla sınırlandırılmasına da olumlu yaklaşmamışlardır. Bu bulgu, ast değerlendirmelerinin, geribildirim ötesinde idari karar verme sürecinde de yer almasının beklendiği şeklinde yorumlanabilir. Katılımcılar ast değerlendirmelerinin ast-üst ilişkisini olumsuz etkileyeceği görüşüne genel olarak fazla katılmamışlardır.

Son olarak, öz değerlendirmelerine katılan katılımcılar her ne kadar kendi yaptıkları öz değerlendirmesinde performanslarını oldukça gerçekçi bir şekilde yansıttıklarını belirttiler de, genel olarak kişilerin kendilerini yansız değerlendirebileceği görüşüne orta düzeyde katılmışlardır. Ayrıca, öz değerlendirmelerinin, özellikle yönetici değerlendirmelerine kıyasla daha “cömertçe” olacağı da kabul gören bir görüştür.

Çalışmanın bulguları bir bütün olarak ele alındığında, geleneksel olmayan değerlendirici gruplarından çalışma arkadaşlarının sağladığı değerlendirmelerin psikometrik kalitesinin çok yüksek olmadığı; kullanıcıların da bu kaynağın kullanılmasına yönelik dikkatli yaklaştıkları görülmektedir. Çalışma arkadaşlarının aynı kişi için yaptıkları değerlendirmeler için hesaplanan korelasyonlara yönelik betimleyici istatistikler incelendiğinde, bazı çalışma arkadaşı gruplarının değerlendirmelerinde aşırı derecede hemfikir oldukları, bazılarının yaptıkları değerlendirmeler arasındaki korelasyonun ise eksi değer taşıdığı görülmüştür. Bu bulgu, çalışma arkadaşı değerlendirmelerinde, değerlendirilen kişinin performansı dışındaki faktörlerin rol oynama olası-

lığının yüksek olduğunu göstermektedir. Bazı grupların çok yüksek düzeyde hemfikir olması diğerlerinin ise son derece düşük düzeylerde hemfikir olması, iş ortamında oluşması muhtemel iç-grup, dış-grup dinamikleri ile açıklanabilir. Bizimki gibi, görece toplulukçu ve ilişki yönelimli kültürlerde, çalışma arkadaşlarının birbirlerini değerlendirmesinde, bireyci kültürlerdekinden farklı süreçlerin devreye gireceği beklenebilir. Çalışmalar, grubun ahenginin bozulması tehlikesi nedeniyle toplulukçu kültürlerde çalışma arkadaşlarının değerlendirme sürecinde yer almasına sıcak bakılmadığını göstermektedir (Entrekin ve Chung, 2001; Huo ve Von Glinow, 1995). Aycan (2000) da, toplulukçu kültürlerde, kayırmacılığın ve sadakatin, çalışma arkadaşlarına ve yöneticilere olumsuz değerlendirme vermeyi engelleyeceğini belirtmektedir.

Bununla beraber, ast değerlendirmelerinin hem psikometrik açıdan hem de kullanıcı tepkileri dikkate alındığında daha kullanılabilir olduğu görülmüştür. Aslında, ulusal kültürel özelliklerimiz düşünüldüğünde astların performans değerlendirme sürecine dahil edilmelerinin problemlili olacağı beklenebilir. Güç mesafesinin bu denli yüksek olduğu (Aycan ve ark., 2000; Hofstede, 1980, 1991) bir kültürde astların değerlendiren pozisyonunda olmasının, hem kendileri hem de değerlendirilen yöneticiler tarafından içselleştirilmesinin oldukça güç bir durum olacağı, bu durumun örgütlerdeki güç dengelerini bozabileceği ileri sürülebilir. Bununla beraber, bu çalışmanın bulguları bu beklentileri desteklememektedir. Astların, çalışma arkadaşlarından farklı olarak, aynı yönetici için görece tutarlı değerlendirmeler sağlayabildiği görülmüştür.

Cömertlik etkisine maruz olduğu bilinen öz değerlendirmelerinin (Thornton, 1980), bu çalışmada da oldukça cömertçe yapıldığı görülmüştür. Bu nedenle de, üst yönetici değerlendirmeleri ile birlikte, öz değerlendirmelerinin herhangi bir değerlendirme sistemi içinde sağlayacağı fayda çok dikkatli değerlendirilmelidir.

Bu çalışmaya yönelik olarak tartışılması gereken önemli bir konu, farklı kaynakların aynı boyutlar üzerinde değerlendirme yapmamış olmalarıdır. Çalışmada, öz, amir, ast ya da çalışma arkadaşı olarak değerlendirme yapan kişiler, daha çok bilgi sahibi olacakları tespit edilmiş olan (yapılan ön çalışmalar sonunda ve de ilgili yazın temelinde) boyutlarda değerlendirme yapmışlardır. Makalenin ön değerlendirmesini yapan değerli bir hakemin de belirttiği gibi, özellikle farklı kaynakların yaptıkları değerlendirmelerin karşılaştırıldığı durumlarda, değerlendirmelerde kullanılan boyutların sayısı ve içerik olarak birbirinden farklı olmasının bir sorun yaratacağı düşünülebilir. Bununla birlikte, bu çalışmada, farklı değerlendirme boyutlarının kullanılmış olmasının yapılan form/kaynak karşılaştırmaları için büyük bir sorun yaratmadığına işaret eden tespitler vardır. Öncelikle, boyut sayıları ve içerikler farklı olsa bile, tüm kaynaklar hem ortamsal hem de görev performansı boyutlarında (farklı sayılarda olsa da) değerlendirme yapmışlardır. Ayrıca, tüm kaynakların yaptığı değerlendirmelerde, değerlendirme örüntüsü boyutlar arasında çok farklılık göstermemiştir. Diğer bir deyişle, yapılan tüm değerlendirmelerde (öz değerlendirmelerinde görece az olmak koşuluyla), hale etkisinin de bir göstergesi olarak, boyutlar arası korelasyonların oldukça yüksek olduğu gözlenmiştir. Tüm bu nedenlerle, farklı kaynakların yaptıkları değerlendirmeleri karşılaştırırken ortalama performans puanının kullanılmış olmasının bu çalışmada ciddi bir problem yaratmadığı söylenebilir.

Bu çalışmanın önemli bir kısıtlılığı, çok kaynaklı değerlendirme sisteminin tam bir uygulamasının yapılamamış olmasıdır. Çalışmada, farklı değerlendirici gruplarından değerlendirmeler alınmış, ancak çok kaynaklı değerlendirmenin en önemli unsuru olan geribildirim, sürece dahil edilememiştir. Sonuç olarak, bu çalışma şekilsel olarak çok kaynaklı değerlendirmenin bazı ölçme problemleriyle birlikte mümkün olabileceğini göstermiştir. Ancak gerçek anlamda çok kaynaklı bir değerlendirme sisteminin işletilebilmesi için, tüm kaynak-

lardan elde edilen değerlendirmelerin yönetici ya da bir danışman tarafından kişiye, performansı geliştirme amaçlı, geribildirim yapılması gerekmektedir. Değerlendirme sürecinde rol oynayacağı ileri sürülen kültürel faktörlerin, çok daha yoğun bir şekilde geribildirim sürecini etkileyeceği söylenebilir. De Luque ve Sommer (2000), performansa yönelik geribildirim arama davranışında etkili olması muhtemel gizil kültürel yapılardan bahsetmektedir. Bu yazarlara göre, örneğin, bizim gibi görece geleneksel kültürlerde doğrudan değil, dolaylı yollardan geribildirim arama ve alma daha çok tercih edilebilmektedir. Türk kültürünün, De Luque ve Sommer'ın bütüncül kültür tanımıyla tutarlı olan ilişki odaklı yönü, yaşam alanları arasında (iş ve özel yaşantı gibi) belirgin ayrımlar yapılmasını engellemektedir. Bu da özellikle olumsuz değerlendirmelerin bireysel algılanması ve iş dışı ilişkilere taşınması, grup ahenginin bozulması ve de çalışanın motivasyonunun azalması ihtimalini artırmaktadır. Davis (1998) de, Türk kültürü gibi ilişki yönelimli kültürlerde insanların özellikle olumsuz performans geribildirimine ani duygusal tepkiler göstermelerinin beklendiği bir davranış olacağını belirtmektedir. Bununla birlikte, vurgulanması gereken önemli bir nokta, kültür temelli bu beklentilerin bir kısmının (belki de çoğunluğunun) Türkiye'de yapılan uygulamalarda desteklenmeyebileceğidir. Nitekim, sunulan çalışmadaki ast değerlendirmelerine yönelik bulgular, var olan kuramsal bilgilerle çok uyumsuzdur. Bu nedenle, çok kaynaklı değerlendirme sistemlerinin işleyip işlemeyeceğini daha sağlıklı değerlendirebilmek için farklı sektörlerde, farklı örneklemeler kullanılarak yapılan ve çok kaynaklı değerlendirme sistemlerinin bütün unsurlarıyla sınındığı görgül çalışmalara ciddi bir şekilde ihtiyaç duyulmaktadır.

Kaynakça

- Antonioni, D. (1996). Designing an effective 360-degree appraisal feedback process. *Organizational Dynamics*, 25, 24-38.
- Atkins, P. W. B., & Wood, R. E. (2002). Self versus others' ratings as predictors of assessment center ratings: Validation evidence for 360-degree feedback programs. *Personnel Psychology*, 55, 871-904.
- Atwater, L., Roush, P., & Fischthal, A. (1995). The influence of upward feedback on self- and follower ratings of leadership. *Personnel Psychology*, 48, 35-59.
- Aycan, Z. (2000). Cross-cultural industrial and organizational psychology: Contributions, past developments, and future directions. *Journal of Cross-Cultural Psychology*, 31(1), 110-128.
- Aycan, Z., Kanungo, R. N., Mendonca, M., Yu, K., Deller, J., Stahl, G., & Kurshid, A. (2000). Impact of culture on human resource management practices: A 10-country comparison. *Applied Psychology: An International Review*, 49(1), 192-221.
- Borman, W. C. (1997). 360° ratings: An analysis of assumptions and a research agenda for evaluating their validity. *Human Resource Management Review*, 7(3), 299-315.
- Bracken, D. W., Timmreck, C. W., Fleenor, J. W., & Summers, L. (2001). 360 feedback from another angle. *Human Resource Management*, 40(1), 3-20.
- Conway, J. M. & Huffcutt, A. I. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance*, 10(4), 331-360.
- Dalessio, A. T. (1998). Using multisource feedback for employee development and personnel decisions. In J. W. Smither (Ed.), *Performance appraisal: State of the art in practice* (pp. 278-330). San Francisco: Jossey-Bass Inc. Publishers.
- Davis, D. D. (1998). International performance measurement and management. In J. W. Smither (Ed.), *Performance appraisal: State of the art in practice* (pp. 95-131). San Francisco: Jossey-Bass Inc. Publishers.
- De Luque, M. F. S., & Sommer, S. M. (2000). The impact of culture on feedback-seeking behavior: An integrated model and propositions. *Academy of Management Review*, 25(4), 829-849.
- Entrekin, L., & Chung, Y. W. (2001). Attitudes towards different sources of executive appraisal: A comparison of Hong Kong Chinese and American managers in Hong Kong. *International Journal of Human Resource Management*, 12(6), 965-987.
- Fletcher, C., & Perry, E. L. (2001). Performance appraisal and feedback: A consideration of national culture and a review of contemporary research and future trends. In N. Anderson, D. S. Ones, H. Kepir-Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work, and organizational psychology*: Vol. 1, pp. 128-144, Sage.

- Frone, M. R., Adams, J., Rice, R. W., & Instone-Noonan, D. (1986). Halo error: A field study comparison of self- and subordinate evaluations of leadership process and leader effectiveness. *Personality & Social Psychology Bulletin*, 12(4), 454-461.
- Garavan, T. N., Morley, M., & Flynn, M. (1997). 360 degree feedback: Its role in employee development. *Journal of Management Development*, 16(2/3), 134.
- Garson, D. (2005). Reliability analysis: Key concepts and terms. www2.chass.ncsu.edu/garson/pa765/reliab.htm#intraclass
- Göregenli, M. (1995). Kültürümüz açısından bireycilik-toplulukçuluk eğilimleri: Bir başlangıç çalışması. *Türk Psikoloji Dergisi*, 10(35), 1-14.
- Göregenli, M. (1997). Individualist-collectivist tendencies in a Turkish sample. *Journal of Cross-Cultural Psychology*, 28(6), 787-794.
- Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology*, 41, 43-62.
- Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Thousands Oaks, CA: Sage.
- Hofstede, G. H. (1991). *Culture's Consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd edition). Thousand Oaks, CA: Sage Publications.
- Huo, Y. P., & Von Glinow, M. A. (1995). On transplanting human resource practices to China: A culture driven approach. *International Journal of Manpower*, 16, 3-15.
- Levy, P. E., & Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management*, 30(6), 881-905.
- London, M., & Beatty, R. W. (1993). 360-degree feedback as a competitive advantage. *Human Resource Management*, 32, 353-372.
- London, M., & Smither, J. W. (1995). Can multi-source feedback change perceptions of goal accomplishment, self-evaluations, and performance-related outcomes? Theory-based applications and directions for research. *Personnel Psychology*, 48, 803-839.
- Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Thousand Oaks, CA: Sage Publications.
- Olmez, A. E., Sumer, H. C., Soysal, M. (2004). Organizational rationality in public, private and multinational firms in Turkey. *Information Knowledge Systems Management*, 4, 107-118.
- Ramamoorthy, N., & Carroll, S. J. (1998). Individualism/collectivism orientations and reactions toward alternative human resource management practices. *Human Relations*, 51, 571-588.
- Seifert, C. F., Yukl, G., & McDonald, R. A. (2003). Effects of multisource feedback and a feedback facilitator on the influence behavior of managers toward subordinates. *Journal of Applied Psychology*, 88(3), 561-569.
- Shrout, P.E., and J. L. Fleiss (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.
- Sözer, S. (2004). *An evaluation of current human resource management practices in the Turkish private sector*. Unpublished manuscript. Middle East Technical University, Ankara, December, 2004.
- Sümer, H. C. (2000). Performans değerlendirmesine tarihsel bir bakış ve kültürel bir yaklaşım. Z. Aycan (Ed.). *Akademisyenler ve profesyoneller gözüyle Türkiye'de yönetim, liderlik ve insan kaynakları uygulamaları*. Ankara: Türk Psikoloji Derneği Yayınları.
- Thornton, G. C., III. (1980). Psychometric properties of self-appraisals of job performance. *Personnel Psychology*, 33, 263-271.
- Triandis, H. C., & Bhawuk, D. P. S. (1997). Culture theory and the meaning of relatedness. In P. C. Earley & M. Erez (Eds.), *New perspectives on international industrial/organizational psychology* (pp. 13-52). San Francisco: Jossey-Bass Inc. Publishers.
- Tsui, A. S., & Ohlott, P. (1988). Multiple assessment of managerial effectiveness: Interrater agreement and consensus in effectiveness models. *Personnel Psychology*, 41, 779-803.
- Westerman, J. W., & Rosse, J. G. (1997). Reducing the threat of rater nonparticipation in 360-degree feedback systems: An exploratory examination of antecedents to participation in upward ratings. *Group & Organization Management*, 22(2), 288-309.
- Yaffee, R. A. (1998). Enhancement of reliability analysis: Application of intraclass correlations with SPSS/Windows v.8. www.nyu.edu/its/socsci/Docs/intracls.html.

Summary

Use of Nontraditional Rating Sources in Performance Appraisals

H. Canan Sümer*

Reyhan Bilgiç

Middle East Technical University

Multi-source feedback systems, or 360 degree feedback systems, have been one of the major developments in the area of performance management over the last two decades (Levy & Williams, 2004). The idea behind multi-source systems is the belief that performance evaluations obtained from multiple sources (i.e., supervisors, peers, subordinates, self, and, if possible, customers) are likely to be more comprehensive and reliable than those provided by a single rating source (i.e., immediate supervisor) (Antonioni, 1996; Bracken, Timmreck, Fleener & Summers, 2001; Dalessio, 1998; London & Smither, 1995). Although multi-source feedback systems are increasingly being used for administrative purposes (Antonioni, 1996), the original idea behind such systems is to contribute to the development of employees through constructive feedback coming from multiple sources.

Performance information available to each rating source is not necessarily the same. While, for example, supervisors are more likely to observe task performance of the ratees, peers are more likely to observe contextual performance. This is why, high correlations between ratings provided by different sources are not expected. Accordingly, studies examining inter-source rating correlations report relatively low correlations between ratings provided by different source pairs (e.g., Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988).

Like many other management theories, models, and applications originating from North America, multi-source feedback systems have been welcome by other countries including Turkey. A recent study on the range and width of human resources management practices in Turkey indicated that there was a positive attitude toward multi-source appraisal in private sector organizations (Sözer, 2004). On the other hand, there are potential problems associated with the application of management systems and theories originating from North America in a totally different cultural context (Aycan, 2000; Aycan, Kanungo, Mendonca, Yu, Deller, Stahl & Kurshid, 2000; De Luque & Sommer, 2000; Fletcher & Perry, 2001; Levy & Williams, 2004; Ramamoorthy & Carroll, 1998; Sümer, 2000). We believe that despite the observed positive attitudes toward multi-source feedback systems in Turkey, we need to evaluate the extent to which such systems fit in our cultural context. Hence, the purpose of this study was to evaluate the feasibility of a multi-source feedback system in a public sector organization and, more specifically, to examine performance appraisals provided by nontraditional rating sources (i.e., subordinates, peers, and self) both psychometrically and on the basis of user-reactions.

*Address for Correspondence: H. Canan Sümer, Orta Doğu Teknik Üniversitesi, Psikoloji Bölümü 06531, Ankara, Turkey.

E-mail: hcanan@metu.edu.tr

Note from authors: Opinions expressed in this manuscript are those of the authors and do not necessarily reflect the views of the organization in which the study was conducted.

Method

Sample

The sample of the study was composed of lower, middle, or upper-middle managers of a public sector organization. The organization in which the study was carried out was a traditional, vertically organized, large public sector organization in which the power distance was very much emphasized. Participants were selected randomly to represent the managerial levels included in the study.

Sample sizes (in terms of the number of raters and ratees) varied depending on the nature of the analyses conducted. However, the basic samples used for each rater-ratee combination were as follows: 251 immediate supervisors¹ (rating 785 subordinates), 29 upper level supervisors (rating 101 subordinates), 302 peers (rating 345 peers, yielding 1638 ratings), 663 subordinates (rating 118 immediate supervisors). Furthermore, 159 managers provided self-ratings.

The age of the participants ranged from 23 to 55 years. In all sub-samples, the males constituted $\geq 90\%$ of the sample.

Measures

Rating forms. Five different rating forms were used in the study (forms used by immediate supervisors, upper level supervisors, self, peers, and subordinates). Forms used by immediate supervisors, upper level supervisors, and self included the same 18 performance dimensions representing both task (e.g., "Quality of Work") and contextual performance (e.g., "Agreeableness") dimensions differing only in instructions. Forms used in peer and subordinate ratings included six and 12 dimensions, respectively. Majority (but not all) of dimensions included in both peer and subordinate rating forms tapped into contextual performance domain. Performance dimensions included in all rating forms were developed

through a multi-stage approach, involving examination of job descriptions and performance evaluation forms in use; focus groups with subject matter experts; administration of questionnaires ($N > 3000$) to potential users to identify relevant dimensions. In all forms, ratings are conducted on a 5-point Likert type scale (1 = Poor Performance; 5 = Exceptional Performance).

User Reactions Questionnaire (URQ). The URQ was filled out by participants providing self (4 items), peer (9 items), and/or subordinate (9 items) ratings to evaluate user reactions toward multi-source performance evaluations in general and toward the use of nontraditional rating sources in particular. Items of the URQ are rated on a 5-point Likert type scale (1 = Strongly Disagree; 5 = Strongly Agree).

Procedure

Participants representing different rating sources provided performance ratings first and then they filled out the URQ. Participants were informed that the evaluations they would provide were to be used for research purposes only. User reactions provided by immediate and upper level supervisors were not included in the present analyses.

Results and Discussion

Consistent with the purpose of the study, two groups of analyses were conducted. First, ratings provided by different rating sources were compared psychometrically. Second, user reactions concerning the use of nontraditional rating sources were analyzed.

Psychometric Qualities of the Ratings Provided by Different Rating Sources

Mean intraclass correlation coefficients (i.e., inter-rater reliabilities) for subordinate evaluations was found to be higher (.49, $sd = .37$, median = .58) than that for peers (.28, $sd = .44$, median = .38), and

¹Consistent with the purpose of this study, in the present analyses only a subset of ratings obtained from immediate supervisors has been used.

also higher than those reported by Conway and Huffcutt (1997) in their meta analysis. Relatively high levels of agreement observed among subordinates rating the same supervisor can be interpreted as an evidence for relative soundness of subordinates as a rating source in the present study. The pattern of correlations between ratings provided by different source pairs (i.e., supervisor-peer = .31, self-peer = .30, subordinate-supervisor = .26, self-supervisor = .17) was similar to those reported in the literature (Atkins & Wood, 2002; Conway & Huffcutt, 1997), with the exception of the correlation between subordinate-peer ratings (.05).

Ratings provided by all sources were prone to leniency. Yet, results of a series of ANOVAs on the average of dimensional ratings indicated that while self and peer ratings were most prone to leniency, immediate supervisors' (not upper supervisors') ratings were relatively less prone to this bias. The observed tendency to provide lenient ratings can be explained by the cultural context in which the system was tried. Fletcher and Perry (2001) refer to studies suggesting a proneness to leniency and an avoidance from negative feedback in cultures characterized by collectivism, such as Turkey (see Hofstede, 1980, 1991).

Finally, analyses of interdimensional correlations indicated that self-ratings were less open to halo than the ratings provided by, especially, peers, upper level supervisors, and subordinates.

Analyses of User Reactions toward Nontraditional Ratings Sources

In general, users were more cautious toward the use of peers in a multi-source performance appraisal system. Although, they were not negative toward the use of peer ratings for feedback (but not for administrative) purposes, they thought that involving peers in performance evaluation could somewhat harm workplace relationships. Compared to peer ratings, users reacted more positively to subordinate evaluations. They were especially positive as long as the rating

subordinate's identity was kept confidential and the ratings were not used for feedback purposes only, suggesting that there was an expectation concerning the use of subordinate evaluations in administrative decision making process as well.

Overall, self-ratings were not perceived to be credible. Participants thought that although they themselves provided objective self-ratings, others would not be able to evaluate themselves objectively.

Taken together, findings of this study suggest that among the nontraditional rating sources, peers seem to be the most problematic one. In addition to problems associated with psychometric qualities, users were not very optimistic about and welcoming towards peer ratings. This finding is consistent with the literature which suggests that peer ratings are not received enthusiastically in collectivistically oriented cultures because of the potential threats against workgroup harmony (Entrekin & Chung, 2001; Huo & Von Glinow, 1995).

Interestingly, however, subordinate ratings were found to be both psychometrically less deficient and have more positive user reactions. This finding may look inconsistent with the Turkish context, which is characterized by relatively high power distance (Aycan et al., 2000; Hofstede, 1980, 1991). One may well argue that involving subordinates in performance evaluation process could be something difficult to accept by both subordinates and supervisors since it is very much against the nature of how things are done in most public sector organizations in Turkey. Yet, findings of this study indicated the opposite. Subordinates do provide more reliable ratings, and users seem more ready for subordinate ratings than they are for peer ratings.

An important limitation of the present study is that it failed to incorporate a test of the feedback component of a multi-source evaluation system. Future studies need to be conducted to fully test the feasibility of a multi-source system in a range of organizations in the Turkish context.