

# Bilişsel Yetenekler Testi'nin (CogAT®) Ana Sınıfı ve Birinci Sınıf Öğrencileri için Kurultu Geçerliliği<sup>1</sup> Çalışması

İ. Ercan Alp \* Ayşesim Diri  
Boğaziçi Üniversitesi

## Özet

CogAT® grup testinin Birincil Bataryası'nın Düzey 1 testi, kurultu geçerliliği açısından değerlendirilmek amacı ile 45 anaokulu ve 360 birinci sınıf öğrencisine ders yılı sonunda uygulanmıştır. Birinci sınıf öğrencileri, üçü özel vakıf ve ikisi de devlet ilköğretim okulu olmak üzere, toplam beş okuldan seçilmiştir. Ana sınıfı öğrencilerinin tümü, yalnızca özel vakıf okullarından birine devam etmekte olan çocuklardır. Çalışmada; (1) birinci sınıf öğrencilerinin puanlarının ana sınıfı öğrencilerinininkilerden ve (2) özel vakıf okullarına devam eden öğrencilerin puanlarının devlet okullarına devam edenlerinkinden daha yüksek olacağı, (3) test puanlarının öğrencilerin okul başarısını, hem eş zamanlı hem de ileri yönelik olarak, güçlü bir biçimde yordayacağı, ve (4) testin yüksek düzeyde içtutarlılığı olacağı beklenmiştir. Tümü ham puanlar üzerinden yapılan analizler bu beklentilerin hepsini doğrulamıştır. Testi oluşturan altı alt-testten yalnızca biri, oral sözcük dağarcığı testi, ile ilgili sonuçlar bu alt-testin bir ölçüde iyileştirilmesinin, testin geçerliliğini daha da yüksek bir düzeye çekeceğine işaret etmektedir. Sonuç olarak bu çalışma, öğretmen ve okul yöneticilerinin yanısıra, alanda çalışan araştırmacılara da yardımcı olacak bir testin kazanılması yönünde atılmış ilk adım olarak görülebilir.

**Anahtar sözcükler:** CogAT®, kurultu geçerliliği, Türk çocukları

## Abstract

To assess its construct validity for Turkish students, the Primary Battery (Level 1) of Cognitive Abilities Test (CogAT®) was administered to 45 kindergarten and 360 grade 1 students late in the Spring term. Three private and two public schools participated in the study. Kindergarten students came only from one of the private schools. In line with the four hypotheses put forward, we specifically predicted that (1) grade 1 students obtain higher scores than kindergarten students, (2) students in private schools obtain higher scores than those in public schools, (3) test scores strongly predict concurrent and later achievement in the school, and (4) the test has a high level of internal consistency. Results of the study confirmed all of these predictions. Only in the case of oral vocabulary subtest, the results suggested a need for further improvement, which would lead to an even higher level of validity for the test. In conclusion, this study has introduced a new test to this culture which would help fill a gap for teachers and school administrators as well as researchers interested in the assessment of cognitive abilities.

**Key words:** CogAT®, construct validity, Turkish pupils

\*Yazışma Adresi: Doç. Dr. İ. Ercan Alp, Boğaziçi Üniversitesi, Psikoloji Bölümü, 34342 Bebek-İstanbul  
E-posta: alpe@boun.edu.tr

<sup>1</sup>"Kurultu geçerliliği" terimi burada "construct validity" karşılığı olarak kullanılmıştır. Her ne kadar aynı amaçla "yapı geçerliliği" terimi yaygın olarak kullanılmakta ise de, kavramı daha iyi ifade ettiğine inandığımız için bu terimi kullanmayı yeğledik.

Yazar Notu: Bu araştırmanın yapılabilmesi için gerekli izinin alınmasında bize yardımcı olan İstanbul İl Millî Eğitim Müdürlüğü yetkililerine, araştırmaya katılan okullarda bize candan yardımcı olan müdür ve yardımcıları, danışman öğretmenler, sınıf öğretmenleri ve tabii ki, testi alan bütün öğrencilere teşekkür etmeyi bir borç biliriz.

Eğitim sistemi içinde öğrencilerin bilişsel yeteneklerinin ölçümü değişik nedenlerden dolayı önem taşır. Öğretmenler ve okul yönetimi açısından, öğrencilerin değişik düzeylerdeki özgül yeteneklerde ulaştıkları düzeyin saptanmasının farklı yararları vardır. Örneğin, ölçüm sonucu belirlenen yetenek profili, öğrencinin o an için göreceli olarak geri ve ileri olduğu yetenekleri ortaya çıkartır ve bu bilgi öğretmeni her öğrenci için, hangi yetenek veya yeteneklerin o öğrencinin yaşına uygun bir düzeye çıkartabilmesi için özel gayret göstermesi gerektiği konusunda aydınlatır. Benzer biçimde, en üst düzeydeki yetenek (genel zeka) puanları, öğrenciler arasında üstün yetenekli ve yaşına göre geri kalmış öğrencilerin olup olmadığı konusunda bir fikir verir ve böyle öğrencilerin daha etraflı bir değerlendirmeye alınmaları gerektiği konusunda uyarır. Okul yönetimi açısından, belli aralıklarla bu tür ölçümlerin yapılması, bilişsel yeteneklerin verilen eğitimden etkileniyor olması nedeniyle okulda uygulanan eğitim programının değerlendirilmesine de yardımcı olabilir. Hatta, özel ilköğretim okullarında öğrenci seçiminde bile bu tür ölçümlerden yararlanılabilir.

Okul başarısını belirlemede etkili olduğu gösterilmiş bilişsel yeteneklerin geniş kapsamlı ölçümü için Stanford-Binet veya WISC-R gibi bireysel zeka testlerinden yararlanılması düşünülebilir. Ancak bu tür uygulamaların geniş çapta yapılması, gerektirdiği zaman ve maddi kaynakların büyüklüğü bakımından gerçekçi değildir. Pratik açıdan grup testlerinin kullanılması kaçınılmaz olmaktadır. Çok az sayıda olan, ana sınıfı yaşındaki çocuklar için de kullanılabilir grup testlerinden ülkemizde standardize edilmiş olanı ise henüz yoktur. Bu çalışma, sözü edilen testlerden Bilişsel Yetenekler Testi (Cognitive Abilities Test, CogAT®) adlı grup testinin standardizasyonu amacı ile başlatmış olduğumuz projenin ilk bölümü olarak gerçekleştirilmiştir.

CogAT®'in son formunu (Form 5) tanıtan, okul yöneticileri için hazırlanmış olan kılavuz (Thorndike ve Hagen, 1996) bu testin, "...okulda ve okul dışında öğrenme ve problem çözme ile ilgili olduğu son 75

yıldır yapılmakta olan görgül çalışmalarla belirlenmiş bir genel ve soyut muhakeme becerileri demetini ..." (s.1) ölçtüğünü öne sürmektedir. Testin birincil amacının okul yaşındaki çocukların (5-18 yaş) sözü geçen bilişsel yeteneklerde eriştikleri düzey ve onların örüntüsünü betimleyici bilgi sağlamak olduğu ve öğrencilerin öğretim hedeflerine ulaşmalarına yardımcı olmak için eğitimlerce bu bilgiden yararlanabileceği ifade edilmiştir. CogAT®'in, Birincil ve Çoklu Düzey Bataryası olmak üzere, farklı yaş grupları için hazırlanmış iki bataryası vardır. Birincil Batarya, ana sınıftan başlayan ve ikinci sınıfın sonuna kadar olan sınıflar için hazırlanmış (hatta 3. sınıfta olup geri kalmış öğrenciler için de kullanılabilir), üç farklı yetenek türünü ölçen ve altı alt-testten oluşan Düzey 1 ve Düzey 2 testlerini içerir. Düzey 1 testi, ana sınıfı ve 1. sınıf için uygundur. Projeimizin bu bölümü, Düzey 1 testinin ülkemizde geçerliğini saptamayı amaçlamaktadır.

CogAT®, geleneksel psikometrik yaklaşım çerçevesinde ortaya çıkmış zeka kuramlarından yararlanarak geliştirilmiş bir testtir. Bu yaklaşım, bireylere farklı tür maddelerden oluşan çok sayıda testin uygulanması ve elde edilen puanların faktör analizi teknikleri kullanılarak irdelenmesi üzerine kurulmuştur (Örn., Embretson ve Schmidt McCollam, 2000). Bu kuramlar arasında zekanın değişik düzeylerde yeteneklerden oluştuğunu savunanlar ile uyumlu olarak CogAT®, genel zeka faktörünün (g) yanısıra, iki alt düzeydeki faktörleri de ölçmeyi hedeflemektedir. Thorndike ve Hagen (1996) CogAT®'in oluşturulmasında özellikle Philip E. Vernon'un hiyerarşik modeli ve Raymond B. Cattell'in akıcı-kristalleşmiş yetenekler (Gf-Gc) modelinden yararlanılmış olduğunu ifade etmektedirler. Testi oluşturan altı alt-testten ilk ikisi, sözel muhakeme ve oral sözcük dağarcığı testleri, birlikte kullanıldığında bir üst düzey yeteneği de (sözel yetenek) ölçmektedir. İlişkisel kavramlar ve nicel kavramlar alt-testleri, benzer şekilde, nicel yeteneği de ölçmektedir. Son iki alt-test, şekil sınıflandırma ve matrisler ise, sözel olmayan yeteneği<sup>2</sup> ölçer. Sonuç olarak testin uygulanması, en alt düzeyde

<sup>2</sup> Testin kılavuzunda bu yetenek "nonverbal ability" diye adlandırılmış olduğu için, "sözel olmayan yetenek" terimini kullandık. Ancak, söz konusu yetenek genelde "görsel-mekansal yetenekler" diye adlandırılan yetenekleri içermektedir.

altı, ara düzeyde üç, ve en genel düzeyde bir yetenek için ayrı ayrı puanların elde edilmesini sağlamaktadır.

Genel zeka faktörü (g) kavramının geçerliği hala ciddi olarak tartışılmaktadır. Örneğin, Gf-Gc kuramının geliştirilmesinde katkıları bulunan John Horn, kuramın almış olduğu son halini anlattığı kitap bölümünde “zeka” değil, “zekalar”ın olduğunu vurgulamakta ve g faktörünün geçersiz olduğunu savunmaktadır (Horn ve Noll, 1997). Kuramın ilk halinde yalnızca iki farklı zeka türü (Gf ve Gc) olduğu öne sürülmüş ise de, daha sonra faktör analizi tekniklerinin kullanıldığı yapısal çalışmaların yanısıra, zihinsel yetenekler ile kalıtım, gelişim, merkezi sinir sistemi, ve başarı arasındaki ilişkilerin incelendiği çalışmaların ışığında, en az yedi farklı zeka türünün olduğu sonucuna varmıştır: muhakeme zekası (reasoning intelligence, Gf), kültürselleşmeden gelen bilgi zekası<sup>3</sup> (the intelligence of knowledge stemming from acculturation, Gc), görsel (Gv), işitsel (Ga), ve nicel zeka (Gq) ile kısa süreli kavrayış ve bellekte tutma (the intelligence of short-term apprehension and retention, SAR) ve akıcılık ve uzun süreli bellekten geriye getirme zeka’sı (the intelligence of fluency and retrieval from long-term storage, TSR). Hatta, zihinsel hız (intellectual speed, Gs) ve karar verme hızı’nın da (speed of decision making, QDS) farklı birer zeka türü olarak görülebilme olasılığında söz etmektedir.

Ancak, tamamen bu konuya adanmış ve alanda bu konu ile uğraşan tanınmış araştırmacıların katkıları ile ortaya çıkmış olan kitabın ilk bölümünde, o kitaptaki yazıların bir değerlendirmesini yapan Demetriou (2002), geleneksel zeka testlerinin g faktörünü ölçüp temsil ettiği konusunda hemen herkesin hemfikir olduğunu dile getirmektedir. Demetriou’ya (2002) göre, g-yandaşları ile g-karşıtları arasındaki fark, g-yandaşlarının, g’nin “mevcut, güçlü ve kullanılabilir her testin ayrılmaz bir parçası” olduğuna, g-karşıtlarının ise “bu durumun, geleneksel zeka

testlerini oluşturan işlemlerin altında yatan varsayımların ve hatta içeriklerinin benzerliğinden kaynaklandığına” inanmalarındır; dolayısıyla karşıtlara göre, “geleneksel zeka testleri eksiktir ve geniş çaplı bir yeniden gözden geçirilme ve düzeltilmelerine gerek duyulmaktadır” (s. 16). Örneğin, Sternberg (2002) bu tür testlerde ölçülen analitik zeka’nın yanısıra, pratik zeka ve yaratıcı zeka’nın da ölçülmesi gerektiğini savunmaktadır. Örnek olarak verdiği görgül çalışmaların sonuçları, pratik ve yaratıcı zeka’yı ölçen testlerin okul-dışı başarının yanısıra, okul başarısını da yordayabileceğini göstermektedir. Ancak, Sternberg’in sözünü ettiği bu iki farklı zeka türünü oluşturan yetenekleri ölçmekte yararlanılabilecek ölçekler içinde yaygın olarak kullanılan herhangi bir test bulunmamaktadır. Onun analitik zeka diye adlandırdığı yetenekler bütünüünün, okul başarısını yordadığı konusunda herhangi bir kuşku yoktur (Örn., Neisser ve ark., 1996). Her ne kadar okul başarısını etkileyen daha başka yetenekler olsa da, yalnızca analitik zeka’nın bile geçerli bir ölçümünün yapılabilmesi bugün için hala önemini korumaktadır.

Analitik zeka diye de adlandırılabilir bilişsel yetenekleri ölçtüğü savunulan CogAT® testinin bu toplumda geçerliğini saptamak amacı ile başlattığımız projenin bu bölümünde yalnızca Düzey 1 testi sınanmıştır. Günümüzde testlerin değerlendirilmesinde kurultu geçerliği ön plana çıkmaktadır. Bir testin hangi kuramsal kurultuyu ne derecede ölçtüğünün belirlenmesi önem kazanmıştır. Kurultu geçerliği, diğer geçerlik türlerinin hepsini içeren en kapsamlı geçerlik olarak da görülebilir (Anastasi ve Urbina, 1997). Geleneksel zeka testlerinin geçerliğinin saptanmasında kullanılmış olan başlıca ölçütlerden birisi gelişimsel değişimdir (Anastasi ve Urbina, 1997). Ham puanların yaş ile artması, o tür testlerin geliştirilmesine ışık tutan zeka kurultusunca yordandığı için, testin geçerli olduğu konusunda bir kanıt olarak görülmüştür. CogAT®’in ölçtüğü yetenekler

<sup>3</sup> Horn ve Noll (1997), kuramın ilk ortaya atıldığı zaman “akıcı zeka” (fluid intelligence, Gf) ve “kristalize/birikimli zeka” (crystallized intelligence, Gc) diye adlandırılmış zekalar için, hala yaygın olarak kullanılmakta olan o terimler yerine, sırası ile, “muhakeme zekası” ve “kültürselleşmeden gelen bilgi zekası” terimlerini kullanmışlardır.

ile geleneksel zeka testlerinin ölçtükleri arasındaki büyük örtüşme nedeni ile, ham puanlar bazında, birinci sınıf çocukları ile ana sınıfı çocukları arasında, birinci sınıf çocukları lehine bir fark olacağı beklenebilir. Bu çalışmada sınıflarımız denencelerden biri budur.

Bu yeteneklerin öğrencinin okulda aldığı eğitimin kalitesi ile artıp azalacağı da, sözü edilen zeka kurultusu ile tamamen uyumludur ve CogAT®'in hazırlayıcılarınca da öne sürülmektedir (Thorndike ve Hagen, 1996). Bu nedenle, bazı özel vakıf ilköğretim okullarında olduğu gibi, küçük sınıflarda, deneyimli öğretmenler tarafından eğitim verilen okullara devam eden çocukların puanlarının, sıradan devlet ilköğretim okullarına devam eden çocuklarınkinden daha yüksek olacağı, çalışmanın bir diğer denencesidir.

Sözü edilen yetenekler, her üç düzeyde de, eğer CogAT® testinin tanıtımında savunulduğu ve oluşturulmasında yararlanılan zeka kurultusunca yordandığı gibi, okul başarısını belirleyen faktörler arasında iseler, o zaman öğrencilerin test puanları ile derslerdeki başarıları arasında anlamlı bir ilişki olmalıdır. Bu da çalışmanın üçüncü bir denencesi olarak ileri sürülmüştür. Bu denencenin bir uzantısı olarak, aynı okulda okumaya devam eden öğrencilerin ileri yıllardaki okul başarısının, testin uygulandığı zaman alınmış olan puan ile anlamlı bir ilişki göstermesi de beklenmiştir.

Son olarak, testlerin içtutarlığının test geçerliği açısından önemli bir yeri vardır çünkü, bir testin belli bir kurultuyu ölçtüğünü gösteren görgül desteğin bulunduğu durumlarda, o testin içtutarlık derecesi, aynı testin uygulanması ile elde edilen puanların o kurultu dışındaki faktörlerden ne ölçüde bağımsız olduğu konusunda bir fikir verir (Anastasi ve Urbina, 1997). Diğer bir deyişle, bir testin içtutarlığı ne denli yüksek ise, o testten alınan puanların, karıştırıcı değişkenlerden o denli az etkilendiği anlaşılır. O nedenle, sınanan dördüncü denence, Düzey 1 testinin,

hem bir bütün olarak hem de her alttest için ayrı, ayrı, yüksek içtutarlık göstereceğidir.

## Yöntem

### Örnekleme

CogAT® Düzey 1 testi, 5 okuldan toplam 405 öğrenciye uygulanmıştır. Bunlardan 45'i ana sınıfı, 360'ı ise birinci sınıf öğrencidir. Bu okullardan üçü özel vakıf okulları, diğer ikisi ise devlet okuludur. Üç özel vakıf okulundan ikisi (Okul A ve B), üst sosyo ekonomik düzeydeki ailelerin çocuklarını gönderdikleri, sınıflardaki öğrenci sayısının düşük tutulduğu ve verilen eğitimin kalitesi konusunda çok duyarlı olduklarından kuşku duyulmayacak okullardır<sup>4</sup>. Test, tek birinci sınıf şubesi olan Okul A'da kayıtlı olan 17 öğrencinin hepsine, Okul B'de üç ana sınıfı şubesinde toplam 45 öğrenciye ve iki birinci sınıf şubesinde toplam 47 öğrenciye uygulanmıştır. Üçüncü özel vakıf okulu (Okul C) ise, verilen eğitimin kalitesi açısından Okul A ve B kadar iddialı olmadığı varsayılanlar arasından seçilmiş, çok sayıda birinci sınıf şubesi olan ve şubelerdeki öğrenci sayısının göreceli olarak biraz daha yüksek tutulduğu bir okuldur. Test, Okul C'de toplam 215 öğrenciye uygulanmıştır. İki devlet ilköğretim okulundan Okul D'de 52 ve Okul E'de 29 öğrenci çalışmaya katılmıştır. Bu okullar, sıradan devlet ilköğretim okullarını temsil etmek üzere seçilmiş ve öğretim elemanları açısından, örnekleme alınmış özel vakıf ilkokullarındaki gibi seçkin bir kadroya sahip olmayan okullardır.

Test uygulamasının her öğrenci için dört ayrı günde ve ders yılının son haftalarında yapılmış olması nedeni ile dört seanstan hepsine katılmış ve altı alttesti de almış öğrenci sayısı, ana sınıfı için 27 ve birinci sınıf için 276'dır. Okul A ve B'deki öğrencilerin cinsiyeti ile ilgili bilgi de kaydedilmiştir. Tüm testi almış ana sınıfı öğrencilerinden, cinsiyet bilgisi alınmış olanların, 9'u kız ve 16'sı erkek; birinci sınıf öğrencilerinin de 26'sı kız ve 29'u erkektir. Okul A ve B olarak ayrıldığında, kız ve erkek sayıları birinci sınıf için, sırasıyla, 9 ve 8, ve 17 ve 21'dir.

<sup>4</sup> Bu okullarda, eğitim kalitesinin kontrolü amacı ile, ölçme-değerlendirme birimlerince her sınıf düzeyindeki öğrencilere düzenli olarak, sonuçları karne notunu etkilemeyen başarı testleri uygulanmakta ve sonuçlar öğretmen ve velilere bildirilmektedir.

### **Veri Toplama Aracı**

CogAT® Düzey 1 ve 2 testleri, öğrencilerden okuma becerileri olması beklenmeksizin uygulanmak üzere hazırlanmış testlerdir. Düzey 1 testini oluşturan altı alt-testten herbiri, 20 adet dört seçeneikli, çoktan seçmeli sorudan oluşur. Düzey 2 testi, ikinci sınıf öğrencileri için hazırlanmış, Düzey 1 testinin her alt-testine fazladan dört madde daha eklenerek genişletilmiş halidir. Her madde için seçenekler birer resim olarak temsil edilmiştir. Aynı satırda yanyana dizilmiş dört resimin altında birer içi boş daire vardır ve öğrencilerden doğru seçeneğin altındaki dairenin içini doldurmaları istenir. Test bütün sınıfa aynı anda uygulanır. Testi uygulayan kişi, her madde için, soruyu yüksek sesle okur ve öğrencilerden doğru cevabı yukarıda belirtildiği gibi işaretlemelerini ister. Örnek olarak, öğrencilerin dikkati (soldan sağa dizilmiş) İngiliz anahtarı, çekiç, kerpeten ve tornavida resimlerinin bulunduğu sıraya çekildikten sonra “Bu sıradaki aletlere bakın. Duvara çivi çakmak için kullanılacak en uygun aletin altındaki daireyi doldurun” der. Herhangi bir zaman kısıtlaması olmadığından, herkesin bir cevap işaretlediğinden emin olduğunda bir sonraki maddeye geçer. Öğrenciler arasında geçmiş deneyimleri nedeni ile test alma becerileri açısından farklar olma olasılığına karşı, bu testi ilk olarak alacak öğrenciler için bir Deneme Testi uygulaması yapılır. Testin uygulaması, tercihen günde bir alt-test veya, günde ikişer alt-test verilerek tamamlanır (Thorndike ve Hagen, 1993). Her alt-test için ham puan, doğru cevaplanmış madde sayısıdır. Orta düzey yetenek bataryalarının herbiri için ham puan, o bataryayı oluşturan iki alt-testten alınmış olan ham puanların toplamıdır. Üç orta düzey yetenek puanlarının toplamı da (başka bir deyişle, altı alt-testten alınan puanların toplamı) bütün testten alınan ham puandır.

CogAT® ile yapılan çalışmaların sonuçlarını özetleyen Thorndike ve Hagen (1996), Birincil Batarya (Düzey 1 ve Düzey 2 testleri) için şu bulgulara yer vermektedir. Testin bütünü için, ders yılının Bahar dönemindeki uygulamada KR20 formülü ile hesaplanmış içtutarlık değeri, ana sınıfı için .945 ve bi-

rinci sınıf için .948 olarak bulunmuştur. Bu değerler, sözel yetenek bataryası için, sırasıyla, .831 ve .832, nicel yetenek bataryası için, .887 ve .880, ve sözel olmayan yetenek bataryası için, .889 ve .914'tür. Toplam puan ve üç orta düzey yetenek puanlarından ilk ikisi ile birinci sınıftaki okuma, matematik ve dil başarısı arasında yüksek düzeyde pozitif korelasyon bulunmuştur. Sözel olmayan yetenek puanı ile başarı arasındaki korelasyon diğerlerine göre biraz daha düşüktür. Testin bütününde, beşi kızlar ve beşi de erkekler için yanlılık gösteren toplam on madde vardır. Altı alt-testin herbirinde kızlar veya erkekler için yanlılık gösteren madde sayısı en fazla üçtür ve bu maddeler iki cinsiyet için yaklaşık olarak dengelenmiştir. Testin bütününden alınan puan (toplam puan) açısından kız ve erkekler arasında herhangi bir fark bulunmamıştır.

Bu çalışma için, Düzey 1 testi önce yazarlarca dilimize tercüme edilmiştir. Toplam 120 maddeden üçü, bu kültüre uygun olmadığı için testten çıkarılmıştır. Hepsisi Amerikan para birimleri ile ilgili olan bu üç madde, nicel kavramlar alt-testinde yer almaktadır. Bu nedenle, çalışmamızda uyguladığımız testin madde sayısı 117'ye inmiştir.

### **İşlem**

Testin uygulamasını yapacak olan asistanlar Psikoloji Bölümü son sınıf öğrencileri arasından seçilmiş ve kendilerine gerekli eğitim verilmiştir. Test önce, Boğaziçi Üniversitesi'nin Okulöncesi Eğitim Merkezi'ne devam etmekte olan ana sınıfı yaşındaki çocuklara uygulanarak bir önsınamadan geçirilmiştir. Bu uygulamayı yapmış olan asistanlar ile yapılan toplantı sonunda, testin uygulanması veya maddelerin anlaşılabilirliği ile ilgili herhangi bir sorun olmadığı ortaya çıkmıştır.

Testin seçilmiş olan beş okulda uygulanması, test kılavuzunda belirtildiği biçimde gerçekleştirilmiştir. Her okulda, birbirini izleyen dört günde, sırası ile, alıştırma testi, birinci ve ikinci alt-testler, üçüncü ve dördüncü alt-testler, ve beşinci ve altıncı alt-testler uygulanmıştır. Uygulamanın dört bölümünün herbiri, bir ders saati içinde tamamlanmıştır. Altı alt-test

aynı kitapçıkta toplanmış olduğundan, öğrencilerden kitapçığın arka kapağına okul numaralarını yazmaları istenmiş ve izleyen günlerde her öğrenciye kendi test kitapçığı geri verilerek testin uygulanmasına devam edilmiştir.

Dört uygulama seansında da, her sınıfta iki asistan hazır bulunmuştur. Asistanlardan biri yönergeyi ve soruları bütün sınıfa yüksek sesle okuyarak testin uygulamasını yaparken diğer asistan sıralar arasında dolaşarak öğrencilerin yönergeye uygun olarak davranıp davranmadıklarını kontrol etmiş, başkalarının cevaplarına bakmalarını engellemek için gayret göstermiş ve gelen sorulardan yalnızca işlemsel olan veya test-alma mekaniği ile ilgili olanları cevaplamıştır.

Çalışmaya katılan bütün okullarda uygulama, ders yılının son iki ayı içinde yapılmıştır. Uygulama-

nın yapıldığı ders yılı sonunda, Okul A'ya devam eden öğrencilerin ders notu ortalamaları ve Okul B'ye devam eden ana sınıfı öğrencilerinin sınıf öğretmeni tarafından yapılmış değerlendirmeleri (5 üzerinden) ile birinci sınıf öğrencilerinin okul yönetimine uygulanmış bir başarı testinden almış oldukları puanlar ikinci veri kaynağıdır. Uygulamayı izleyen üçüncü yıl sonunda her üç özel vakıf okuluna hala devam etmekte olan öğrencilerin o yılki not ortalaması da üçüncü veri kaynağı olmuştur. Bu veriler, öğrencilerin adları saklı kalarak yalnızca okul numaraları bazında elde edilmiştir.

### Bulgular

Bu bölümde önce, test hakkında genel bir fikir vermek amacı ile okul bazında puan ortalamalarının yanı sıra, testi oluşturan maddelerin zorluk düzeyleri ile ilgili bulgulara, ve kız ve erkek öğrencilerin puan-

**Tablo 1**

*Okul ve Sınıf Düzeyine Göre Testten Alınan Puanların Ortalama ve Standart Kayma Değerleri*

Okul	Puan Ortalamaları						
	Sözel Muhakeme	Oral sözcük dağarcığı	İlişkisel kavramlar	Nicel kavramlar	Şekil sınıflandırma	Matrisler	Toplam
A	13.29 (3.51)	13.24 (2.66)	16.47 (1.91)	13.88 (2.76)	14.76 (4.44)	15.35 (4.83)	87.00 (15.72)
	17	17	17	17	17	17	17
B							
Ana sınıfı	10.27 (3.65)	11.51 (3.09)	13.53 (3.64)	8.86 (4.08)	10.94 (3.22)	10.24 (4.83)	64.63 (17.17)
	45	45	36	36	33	34	27
Birinci sınıf	15.23 (2.62)	13.64 (2.08)	16.33 (3.17)	14.79 (2.94)	16.64 (2.84)	17.72 (3.15)	95.08 (11.35)
	47	47	43	43	42	43	38
C	12.18 (3.29)	12.67 (2.48)	15.36 (2.90)	12.71 (3.42)	13.79 (4.22)	13.97 (5.21)	80.35 (16.98)
	215	215	203	203	205	205	194
D	7.46 (3.14)	10.35 (2.31)	10.81 (2.82)	5.33 (2.54)	9.64 (2.71)	6.88 (3.05)	49.00 (7.86)
	52	52	21	21	25	25	11
E	10.07 (3.41)	12.17 (2.55)	14.40 (2.74)	11.90 (3.57)	12.76 (3.63)	10.10 (4.58)	72.19 (14.38)
	29	29	20	20	21	21	16

Not. Puanların standart kayması, parantez içinde verilmektedir. Ortalamaya dahil edilmiş öğrenci sayısı, standart sapma değerlerinin altında yer almaktadır.

larının karşılaştırılmasının sonucuna yer verilmektedir. Daha sonra, doğrulanması halinde testin kurultu geçerliğinin yüksek düzeyde olduğunu gösterecek, araştırma öncesi kurulmuş denenceleri sınamak üzere yapılmış analizlerin sonuçları sunulmaktadır. Bütün analizler ham puanlar üzerinden yapılmıştır.

### Ön analizler

Çalışmaya katılmış olan okullardaki öğrencilerin puan ortalamaları, standart kayma değerleri ile birlikte, testin tümü ve altı alt-test için ayrı olarak Tablo 1’de verilmektedir.

Düzey 1 testini oluşturan alt-testlerdeki maddelerin herbiri için zorluk düzeyi, doğru cevap yüzdesi olarak hesaplanmıştır. Her alt-test için, o testi oluşturan maddelerin zorluk düzeyi ortalaması ve aralığı ana sınıfı ve birinci sınıf öğrencileri için ayrı olarak Tablo 2’de verilmektedir. Tablo 2’de görüldüğü gibi, maddelerin zorluk düzeyi genelde geniş sayılabilecek bir aralık arasında değişmektedir ve her iki sınıf düzeyinde de öğrencilerin hepsinin doğru olarak cevapladığı bir madde yoktur.

**Tablo 2**

*Testin Bütünü ve Alt-testlerin Zorluk Düzeyi Değerleri*

Test	Ana Ana sınıfı	Birinci sınıf
Sözel muhakeme	51.3 (20.0-73.3)	58.9 (20.6-86.7)
Oral sözcük dağarcığı	57.6 (20.0-93.3)	62.2 (12.5-96.1)
İlişkisel kavramlar	67.6 (19.4-97.2)	75.9 (19.7-99.0)
Nicel kavramlar	52.1 (8.3-97.2)	73.6 (33.9-96.4)
Şekil sınıflandırma	54.7 (27.3-93.9)	69.1 (43.5-90.6)
Matrisler	51.2 (23.5-91.2)	68.7 (45.0-97.1)
Bütün test	56.2 (8.3-97.2)	67.9 (12.5-99.0)

Not. Zorluk düzeyi, doğru cevap yüzdesi olarak ifade edilmiştir. Her test için zorluk düzeyi aralığı parantez içinde verilmektedir.

Test puanlarının cinsiyete göre farklılık gösteriyor olma olasılığına karşı, cinsiyet bilgileri elimizde olan Okul A ve B öğrencileri için, kız ve erkek öğrencilerin toplam test puanları karşılaştırılmış ve anlamlı bir fark bulunmamıştır,  $t = .94$ ,  $sd = 78$ ,  $p = .35$ . Birinci sınıf ve ana sınıfı olarak ayrı, ayrı karşılaştırıldığında da, kız ve erkekler arasında herhangi bir fark bulunmamıştır; sırasıyla,  $t = .81$ ,  $sd = 53$ ,  $p = .42$  ve  $t = .28$ ,  $sd = 23$ ,  $p = .78$ . Testin cinsiyet açısından yanlılık göstermediği anlaşılmaktadır.

**Testin içtutarlığı**

Altı alt-test için tüm örneklem üzerinden bulunmuş olan içtutarlık katsayıları

(KD20 fomülü ile hesaplanmış olarak) Tablo 3’de verilmiştir. Alt-test bazında saptanmış olan içtutarlık katsayıları arasında en düşük değer oral sözcük dağarcığı testi için (.55), en yüksek değer de matrisler testi için (.91) bulunmuştur. Testin bütünü (.97) ve sözel, nicel, ve sözel olmayan yetenekleri ölçen bataryalar için bu çalışmada bulunan değerler (sırasıyla, .79, .97, ve .97), genel olarak testin çok yüksek düzeyde içtutarlığı olduğunu göstermektedir. Bu sonuçlar sınanmakta olan denenceyi desteklemektedir.

**Tablo 3**

*Tüm Örneklem Üzerinden Testin İçtutarlık Katsayıları*

Test	KR20	n
Sözel muhakeme	.73	405
Oral sözcük dağarcığı	.55	405
İlişkisel kavramlar	.75	340
Nicel kavramlar	.86	340
Şekil sınıflandırma	.81	343
Matrisler	.91	345
Sözel batarya	.79	405
Nicel batarya	.97	340
Sözel olmayan batarya	.97	343
Bütün test	.97	303

### Okulların Karşılaştırılması

Çalışmaya katılmış okullar arasındaki farklar ile ilgili denenceyi sınamak için okullar önce, özel vakıf ve devlet okulları olmak üzere iki gruba ayrılmıştır. Üçü de özel vakıf okulu olan Okul A, B, ve C’ye ve ikisi de devlet okulu olan Okul D ve E’ye devam eden birinci sınıf öğrencilerinin toplam puan ortalamaları, sırasıyla, 83.05 ( $n = 249$ ) ve 62.74 ( $n = 27$ )

olarak bulunmuştur. Aradaki fark anlamlıdır,  $t = 5.91$ ,  $sd = 274$ ,  $p < .001$ . Orta düzey yetenekler açısından yapılan karşılaştırmalar o düzeyde de özel vakıf okulları lehine anlamlı farklar olduğunu göstermiştir. İki grubun sözel, nicel ve sözel olmayan yetenek puanları karşılaştırıldığında bulunan değerler, sırasıyla,  $t = 9.53$ ,  $sd = 358$ ,  $p < .001$ ;  $t = 6.44$ ,  $sd = 48.63$ ,  $p < .001$ ; ve  $t = 8.41$ ,  $sd = 71.92$ ,  $p < .001$  olmuştur<sup>5</sup>. En alt düzeydeki yetenekleri temsil eden altı alt-test puanları için bulunan  $t$  değerlerinin hepsi 5.63 veya daha yüksek, ve  $p$  değerlerinin hepsi .001'den düşüktür. Bu sonuçlar, özel vakıf okullarına devam eden öğrencilerin testten aldıkları puanların devlet okullarına devam eden öğrencilerin puanlarından daha yüksek olduğunu göstererek sınamadaki denenceyi desteklemiştir.

Tablo 1'de görüldüğü gibi, bu iki grupta yer alan okulların puan ortalamaları kendi aralarında da farklılık göstermektedir. Aynı grup içindeki okullar arasındaki farkların anlamlı olup olmadığını saptamak amacı ile ikinci bir dizi istatistiksel analizler yapılmıştır. Hepsi özel vakıf okulu olan Okul A, B, ve C'ye devam eden öğrencilerin testten aldıkları toplam puanlar arasında anlamlı bir farklılık olduğu bulunmuştur, ( $F(2, 246) = 13.73$ ,  $p < .001$ ). Bu farklılığın hangi okullar arasındaki farklardan kaynaklandığını belirlemek için, varyansların homojen olmadığını gösteren Levene testi sonucundan ötürü, Dunnett C testi uygulanmıştır. Bu testin sonuçlarına göre, Okul B öğrencilerinin toplam puanlarının Okul C öğrencilerinininkinden daha yüksek olduğu (ortalama fark = 14.73,  $p < .05$ ) ve Okul A öğrencilerinin toplam puanları ile Okul B ve Okul C öğrencilerinin toplam puanları arasında herhangi bir fark olmadığı belirlenmiştir. Orta düzey yetenek puanlarının karşılaştırmalarından çıkan sonuçlar da aynı yöndedir. Tek yönlü varyans analizlerinden çıkan değerler, sözel, nicel ve sözel olmayan yetenek puanları için, sırasıyla,  $F(2, 276) = 13.09$ ,  $p < .001$ ,  $F(2, 260) = 5.65$ ,  $p < .005$ , ve  $F(2, 261) = 11.37$ ,  $p < .001$ , olmuştur. Okulların kendi aralarındaki karşılaştırmalarının yapılma-

sı için kullanılmış olan, yeteneklerden ilk ikisi için Scheffe ve üçüncüsü için Dunnett C testi, üç yetenek için de Okul B puan ortalamalarının Okul C'den daha yüksek (ortalama farklar: sırasıyla, 4.03, 3.04 ve 6.58;  $p$ 'ler  $< .05$  veya daha düşük) ve Okul A puan ortalamalarının diğer iki okuldan farksız olduğunu göstermiştir. Aynı analizler altı alt-test puanları ile yinlendiğinde, yalnızca ilişkisel kavramlar alt-testi dışında,  $F(2, 260) = 2.82$ ,  $p = .061$ , bütün alt-test puanları için de aynı sonuçların geçerli olduğu anlaşılmıştır ( $F$  değerleri 3.27 veya daha yüksek,  $p$  değerleri .05 veya daha düşük; Okul B ve C arasındaki ortalama farklar .97 veya daha yüksek,  $p$  değerleri .05 veya daha düşük). Bu üç okul arasında *ilişkisel kavramlar* alt-testi puanları açısından ise, anlamlı bir fark ortaya çıkmamıştır. Genel olarak, üç yetenek düzeyinde de en yüksek puan ortalamaları Okul B ve en düşük puan ortalamaları Okul C öğrencilerindedir. Okul A öğrencilerinin puan ortalamaları Okul B'den düşük ve Okul C'den yüksek olmakla birlikte, bu farklar istatistiksel açıdan anlamlı değildir.

İkisi de devlet okulu olan Okul D ve E öğrencileri karşılaştırıldığında, Okul E öğrencilerinin toplam puanlarının daha yüksek olduğu bulunmuştur;  $t = 4.85$ ,  $sd = 25$ ,  $p < .001$ . Okul E öğrencilerinin üç orta düzey ve altı en alt-düzyer yetenek puanlarının hepsi tutarlı bir biçimde Okul D öğrencilerinininkilerden daha yüksek bulunmuştur. Hesaplanan  $t$  değerlerinin hepsi 2.75 veya daha yüksek ve  $p$  değerleri .01 veya daha düşüktür.

Özet olarak, özel vakıf ve devlet okullarından örneklemimize katılmış olanlar kendi aralarında farklılık gösteriyor olsalar da, grup olarak özel vakıf okullarının puan ortalamaları devlet okullarının puan ortalamalarından daha yüksek olarak bulunmuştur. Ayrıca, verilen eğitimin kalitesi açısından diğer iki özel vakıf okulundan göreceli olarak daha düşük olduğu izlenimini edindiğimiz Okul C'nin puan ortalamaları, üç yetenek düzeyinin her birinde, bir alt-test dışında, Okul B puan ortalamalarından daha düşük olarak bulunmuştur. Bu sonuçlar, okulda verilen eğitimin

<sup>5</sup> İki grubun nicel ve sözel olmayan yetenek bataryalarından aldıkları puanların varyanslarının, Levene testi ile karşılaştırıldığında eşit oldukları anlaşıldığından, o puanların karşılaştırmalarının yapılmasında, varyansların eşit olmadığını varsayan  $t$  testi kullanılmıştır.



kalitesi ile test puanları arasında bir ilişki olduğunu öne süren denenceyi desteklemiştir.

### ***Ana sınıfı ve birinci sınıf öğrencilerinin karşılaştırılması***

Sınanmakta olan denencelerden bir diğeri de, ham puanların yaş ile artacağıdır. Birinci sınıf öğrencileri, grup olarak, ana sınıfı öğrencilerinden yaş olarak daha büyük oldukları için, test puanlarının da daha yüksek olacağı beklenmiştir. Çalışmaya katılan okullar arasında test puanları açısından anlamlı farklar bulunduğu için bu denence, ana sınıfı öğrencilerini yalnızca kendi okullarındaki birinci sınıf öğrencileri ile karşılaştırarak sınanmıştır. Okul B'deki ana sınıfı ve birinci sınıf öğrencileri arasında, üç düzeydeki yetenek puanlarının herbiri için beklenen yönde anlamlı farklar olduğu görülmüştür. Birinci sınıf öğrencilerinin toplam puanı ana sınıfı öğrencilerinden yüksektir,  $t = 8.61$ ,  $sd = 63$ ,  $p < .001$ . Orta düzey yetenek puanları da daha yüksek olarak bulunmuştur: sözel, nicel<sup>6</sup>, ve sözel olmayan yetenek puanlarının karşılaştırmalarında bulunan değerler, sırasıyla,  $t = 6.63$ ,  $sd = 90$   $p < .001$ ;  $t = 5.75$ ,  $sd = 76$ ,  $p < .001$ , ve  $t = 9.41$ ,  $sd = 73$ ,  $p < .001$  olmuştur. Her alt-test için ayrı olarak bakıldığında da, bulunan t değerlerinin hepsi 3.65 veya daha yüksek olup  $p < .001$  düzeyinde anlamlıdır. Bu sonuçlar, Düzey 1 testinde puanların yaşla arttığını, dolayısıyla, yaş ayırımı ölçütü ile bakıldığında testin geçerli olduğunu göstermektedir.

### ***Test Puanları ile Okul Başarısı Arasındaki İlişki***

Test puanlarının okul başarısını yordayacağı yönündeki denence hem eş zamanlı, hem de ileriye yönelik olmak üzere iki ayrı açıdan sınanmıştır. Eş zamanlı yordayıcılığın belirlenmesinde yalnızca Okul A ve Okul B öğrencilerinden toplanan verilerden ya-

rarlanılmıştır. Okul C yönetimince bize söylendiği üzere, birinci sınıf düzeyinde verilen notların başarı ile ilişkisinin zayıf olması nedeni ile (hemen her öğrenci bütün derslerden tam not almaktadır), o okul öğrencileri için bu karşılaştırma yapılmamıştır. Aynı durum Okul B için de geçerli olduğu halde, o okulda tüm öğrencilere okul yönetimince bir başarı testi uygulanmış olduğundan, o testten alınan puanlardan başarı ölçütü olarak yararlanılmıştır. Yine Okul B'de ana sınıfı öğretmenlerinden, ders yılı sonunda öğrencilerin başarı düzeyini 5 üzerinden değerlendirmeleri istenmektedir. Sınıf öğretmenlerince yapılmış bu değerlendirmelerin sonuçları da ana sınıfı öğrencileri için başarı ölçütü olarak kullanılmıştır.

Okul A öğrencilerinin test puanları ile ders notu ortalamaları (Hayat Bilgisi, Türkçe, ve Matematik notlarının ortalaması) ile toplam puan arasında anlamlı bir ilişki bulunmuştur<sup>7</sup>:  $r(16) = .63$ ;  $p < .02$ . Okul B ana sınıfı öğrencilerinin, sınıf öğretmenlerince herhangi bir sınav yapmaksızın kendi izlenimleri bazında verdikleri değerlendirme puanı ile toplam test puanı arasında da anlamlı bir ilişki bulunmuştur,  $r(27) = .66$ ,  $p < .001$ . Aynı okuldaki birinci sınıf öğrencilerinin, okul başarısını ölçmek için okul yönetimince verilmiş olan başarı testinden aldıkları ortalama puan ile toplam test puanı arasında bulunan ilişki de anlamlıdır<sup>8</sup>:  $r(37) = .68$ ;  $p < .001$ . Bu sonuçlar testin yüksek düzeyde eş zamanlı yordayıcı geçerliliği olduğunu göstermektedir.

Testin ileriye yönelik yordayıcılığının sınanması için, üç özel vakıf okulunda testi birinci sınıfta iken almış ve hala aynı okula devam etmekte olan öğrencilerin toplam test puanı ile 3 yıl sonraki ders notu ortalamaları (Türkçe, Matematik, Fen, ve Sosyal Bilgiler dersleri üzerinden) arasındaki ilişkiye bakılmış ve bu ilişkinin anlamlı olduğu bulunmuştur:  $r(177)$

<sup>6</sup> Nicel yetenek bataryasını oluşturan ilişkisel kavramlar ve nicel kavramlar alt-testlerinin ikisini de almış ana sınıfı öğrencisi sayısı 35'tir. Bu alt-testleri almış olan toplam ana sınıfı öğrencisi sayısı, herbiri için 36 olsa da, bir öğrenci ilişkisel kavramlar alt-testini almış olduğu halde nicel kavramlar alt-testini almamış, bir diğerinde de tam tersi olmuştur. O nedenle, nicel yetenek puanlarının karşılaştırması 78 denek üzerinden yapılmıştır.

<sup>7</sup> Okul A öğrencilerinden biri, ders yılı sona ermeden hemen önce okuldan ayrıldığı için, o öğrencinin ders yılı sonu notlarına ulaşamamıştır. Bu nedenle, test puanları ile başarı arasındaki korelasyon katsayısı 16 denek üzerinden hesaplanmıştır.

<sup>8</sup> Okul B birinci sınıf öğrencilerinden birinin başarı testi sonuçları elimize geçmemiş olduğu için bu korelasyon, 37 öğrenciden elde edilmiş veriler üzerinden hesaplanmıştır.

= .50,  $p < .001$ . Örnekleminizdeki ana sınıfı öğrencilerinin toplam test puanları ile üç yıl sonra aldıkları ders notu ortalaması (Türkçe, Matematik, ve Hayat Bilgisi üzerinden) arasında da anlamlı bir ilişki bulunmuştur: ( $r(14) = .65, p < .02$ ).

Okullar arasında ders değerlendirmede kullanılan standartlar açısından farklılıklar olabileceği düşünülerek, birinci sınıf katılımcıları için aynı analizler üç okulun herbiri için tekrarlanmıştır. Bu yöndeki kuşku- ları doğrular biçimde, farklı katsayılar bulunmuştur. Toplam puan ile okul başarısı arasındaki en yüksek katsayı Okul A için,  $r(9) = .95, p < .001$  ve en düşük katsayı Okul C için,  $r(138) = .50, p < .001$ , bulunmuştur. Okul B (yalnızca testi birinci sınıfta almış öğrenciler) için bulunan değer, diğer ikisinin arasındadır:  $r(30) = .61, p < .001$ . Bulunan katsayılar arasında ciddi nicel farklılıklar olsa da, üç okulun herbiri için bulunan katsayıların hepsi en azından orta düzeyde güçlü bir ilişkiye işaret etmekte ve istatistiksel açıdan çok yüksek düzeyde anlamlıdır. Dolayısıyla bu bulgular, testin ileriye yönelik yordalayıcı geçerliğinin de olduğunu göstermektedir. Ders notlarının verilmesinde, bir ölçüde özneliğin olması kaçınılmazdır ve öğretmenler arasındaki olası farklılıkların bu katsayıları olumsuz yönde etkileyeceği beklenebilir. O nedenle, okul başarısının nesnel başarı testleri ile ölçülmesi durumunda elde edilecek değerlerin burada sunulan katsayılardan da daha yüksek olacağı kanısındayız.

### Tartışma

Bu çalışmada, CogAT® Düzey 1 testinin Türkçe'ye tercüme edilmiş formunun geçerliğinin belirlenmesi için dört denence sınanmıştır. Bu denencelerin herbiri için destekleyici kanıt bulunmuştur. İlk olarak, öngörülmüş olduğu gibi, (1) özel vakıf okullarına devam eden öğrencilerin test puanlarının, devlet okullarına devam eden öğrencilerinkinden ve (2) birinci sınıf öğrencilerinin test puanlarının, grup olarak kendilerinden yaşça küçük ana sınıfı öğrencilerinden daha yüksek olduğu bulunmuştur. Dolayısıyla, testin oluşturulmasına temel olarak alınmış zeka kurultusu ile uyumlu olarak, puanların yaşla arttığı ve okulda verilen eğitimin kalitesi ile doğrudan iliş-

kili olduğu görgül olarak saptanmıştır. Özel vakıf okulu olmasına karşın, verilen eğitimin kalitesi açısından diğer iki özel vakıf okulundan (Okul A ve B) biraz daha düşük olduğu izlenimini edindiğimiz Okul C için bulunan farklı düzeylerdeki yetenek puanlarından hemen hepsinin Okul B puanlarından düşük olması da, okulda verilen eğitim kalitesi ile test puanları arasındaki ilişki üzerine kurulmuş olan denencemiz ile tutarlıdır.

İkinci olarak, test puanlarının hem eş zamanlı ve hem de üç yıl sonra ölçülmüş okul başarısı ile oldukça güçlü bir ilişkisi olduğu bulunmuştur. Bu genel bulgu, denencemizi destekleyerek testin yordayıcı geçerliğinin var olduğunu göstermiş ve testin dayandığı zeka kurultusu ile de uyumlu olduğundan, testin kurultu geçerliği hakkında destekleyici kanıt olmuştur. Son olarak, testin içtutarlılığının çok yüksek bir düzeyde olduğu bulunmuştur.

Çalışmamızın sonuçları, okulda verilen eğitimin kalitesi ile testten alınan puanlar arasında bir ilişki olduğu denencesini desteklemiş olmakla birlikte, örnekleme katılmış okulların seçiminde nesnel bir ölçüt kullanılmamış olması bu çalışmanın zayıf noktalarından biri olarak görülebilir. Her ne kadar okul seçiminde, sınıf büyüklüğü, öğretmenlerin titizlikle seçimi, okulda verilen eğitimin denetlenmesi için bir ölçme-değerlendirme biriminin olup olmadığı ve eğer varsa, o birimin ne denli etkin olduğu gibi değişkenlere dikkat edilmiş ise de, bu ölçütlere göre ayrıntılı nicel ölçümler yapılmamış, genel izlenim üzerinden karar verilmiştir. Söz konusu denencenin, bu çalışmada olduğu gibi kabaca değil, güçlü olarak sınanması için bu değişkenin nesnel olarak ölçülmesi ve örnekleme katılacak okulların o şekilde belirlenmesi gerekir. Böyle bir yaklaşım ile, devlet ve özel vakıf okullarının arasındaki farkın, okulda verilen eğitimin kalitesinden kaynaklandığı ve bu iki tür okulların kendi aralarındaki farkların da yine eğitim kalitesi ile ilintili olduğu, kuşku bırakmayacak bir biçimde gösterilebilir.

Bu çalışmaya katılan öğrencilerin ailelerinin sosyoekonomik düzeyi hakkında bireysel bazda veri toplanmamıştır. Her ne kadar çocukları özel vakıf

ilköğretim okullarına devam eden ailelerin büyük çoğunluğu yüksek sosyoekonomik düzeyde olsalar da, o okullarda burslu olarak okuyan çocuklar daha alt düzey ailelerin çocuklarıdır. Benzer biçimde, devlet ilköğretim okullarına devam eden öğrenciler arasında orta sosyoekonomik düzey ailelerin çocukları da olabilir. Ayrıca ülkemizde son zamanlarda gözlenmekte olan eğitim ile gelir arasındaki aşırı denge bozukluğu nedeni ile, sosyoekonomik düzeyin hangi ölçütlere göre belirlenmesi gerektiği de pek açık değildir. Bu durum, batı toplumlarında bulunmuş bilişsel yetenekler ile sosyoekonomik düzey arasındaki ilişkilerin, ülkemiz için ne denli genelleştirilebileceği sorusunu da beraberinde getirmektedir. Bu çekincelerin ötesinde, bu çalışma çerçevesinde sosyoekonomik düzey değişkeni ile okulda verilen eğitimin kalitesi değişkeninin ayrıştırılması olanaksızdır. Bu da çalışmamızın zayıf noktalarından biri diğeri olarak görülebilir. Ancak, veri toplamak için gidilmiş okulların, verdikleri eğitimin kalitesi açısından ciddi farklılıklar gösterenler arasından seçilmesine gayret edilmiştir. Ayrıca CogAT® testi, özellikle okulda verilen eğitimden etkilendiği bilinen yetenekleri ölçmek üzere hazırlanmış bir testtir. Dolayısıyla, okullar arasında bulduğumuz farkların, hiç olmazsa bir ölçüde, o okullarda verilen eğitimin kalitesinden kaynaklandığı çıkarımını yapmak yanlış olmayacaktır. Özellikle testin ders yılı sonunda uygulanmış olması, okulda verilen eğitimin etkisinin en yüksek düzeye çıkarılmış olduğunu düşündürmektedir.

Çalışmamızın bir diğer zayıf noktası, yalnızca bir okulun ana sınıfı öğrencilerinden veri toplayabilmiş olmamızdır. O nedenle, hem örnekleminizin ana sınıfı öğrencisi bölümü sayıca çok düşük kalmış hem de, okullar arasında karşılaştırma olanağından yoksun kalmıştır. Dolayısıyla, ana sınıfı ile ilgili sonuçların genelliği hakkında bir çıkarım yapmamız zorlaşmıştır. Yine de, ana sınıfı öğrencilerinden toplanan veriler ile birinci sınıf öğrencilerden toplanan verilerin analizlerinden çıkan sonuçlar birbirleri ile tamamen uyumludurlar ve denencelerimizi desteklemişlerdir. O nedenle, genelleştirilmeleri ile ilgili ciddi bir kuşkuumuz yoktur. Ancak, bulgularımızın bun-

dan sonra yapılacak çalışmalar ile doğrulanması (replication) gerektiği de açıktır.

Bu çalışmada testin uygulamasının ders yılı sonunda yapılması bazı nedenlerle tercih edilmiştir. Önce, yukarıda değinildiği gibi, böylece okulun öğrencilerin yetenekleri üzerindeki etkisinin en yüksek düzeyde olması sağlanmıştır. Ayrıca, her ne kadar bu test okuma becerileri gerektirmiyorsa da, verilerin sağlığı açısından öğrencilerin okul ortamına uyum sağlamış olmaları gerekmektedir. Ana sınıfına devam etmeden birinci sınıfa başlamış öğrencilerin de olabileceği düşünülerek her iki sınıf düzeyinde de uygulama ders yılı sonunda yapılmıştır. Ancak testin kılavuzunda, birinci sınıfın sonunda, zorluk derecesi daha yüksek olan Düzey 2 testinin kullanılması önerilmektedir. Biz ise, sınıflar arası karşılaştırmayı kolaylaştırmak amacı ile, her iki sınıf düzeyinde de Düzey 1 testini uygulamayı yeğledik. Uygulamanın ders yılı sonunda yapılmış olması nedeniyle, bu çalışmada birinci sınıf öğrencileri için, testin zorluk düzeyi göreceli olarak azaltılmış olmuştur. Yine de, testin okul başarısını yordalayıcı gücünün yüksek olduğunu gösteren sonuçlar, bu durumun herhangi bir sorun yaratmadığını kanıtlamaktadır.

Testin uygulandığı zaman ile ilgili olarak bir sorun, ders yılı sonunda okula gelmeme sıklığının yüksek oluşudur. Uygulama için aynı asistan ekibi her hafta bir okula gitmiştir. Ders yılının son haftalarına gelindiğinde gidilmiş okullarda testin uygulandığı günlerden bir veya daha fazlasına gelmeyen öğrenci sayısında artış olduğu gözlenmiştir. Bu nedenle testten alınan toplam puanlar ile yapılan analizlerde kullanılan denek sayısında ciddi azalmalar olmuştur. Örneğin en son gidilmiş okullardan Okul D'de, örneklem büyüklüğü 52'den 11'e düşmüştür. Bundan sonra yapılacak çalışmalarda uygulamanın bu denli geç kalmamasına dikkat edilmesi doğru olacaktır.

Değiniilmesi gereken bir diğer nokta da, çalışmaya katılmış olan öğrencilerin cinsiyeti ile ilgili bilginin yalnızca iki özel vakıf okulu öğrencilerinden alınmış olmasıdır. Her ne kadar, kız ve erkek öğrencilerin toplam test puanları arasında anlamlı bir fark

bulunmamış ise de, o iki okula devam eden öğrencilerin büyük çoğunluğu yüksek sosyoekonomik düzeydeki ailelerin çocuklarıdır. Aynı sonucun diğer, ve özellikle en alt, sosyoekonomik düzeydeki ailelerin çocukları için de geçerli olup olmadığının görgül olarak saptanması, test puanlarının cinsiyet açısından yanlılık göstermediğine dair bulgumuzun ne denli genel olduğu konusuna açıklık getirmesi açısından gereklidir.

Ancak bu çalışmada, testin eşzamanlı ve ileriye yönelik yordayıcı gücü ile ilgili olarak bulunan sonuçlar, yukarıda değinilen, çalışmanın zayıf kaldığı noktaların önemini bir ölçüde azaltmaktadır. Örneğin, üç ayrı okulda 177 birinci sınıf öğrencisinden toplanmış veriler üzerinden hesaplanmış, test puanlarının üç yıl sonra, nesnellik açısından çok güvenilir olması beklenmeyen, öğretmen değerlendirmesi sonucu verilmiş notları .50 düzeyinde yordadığını gösteren sonuç çok etkileyicidir. Bu değer, geleneksel zeka testlerinin uzun yıllar yaygın olarak uygulanmış ve araştırılmış olduğu ABD’de yapılmış görgül çalışmaların sonuçlarını özetleyen raporda, zeka testi puanları ile okulda alınan notlar arasındaki korelasyon için verilen yaklaşık değerle aynıdır (Neisser ve ark., 1996, s. 81).

Genel olarak bulguların hep beklenen yönde ve denenceleri güçlü bir biçimde destekleyici olmasına karşın, testin daha da geçerli bir hale getirilebileceği de açıktır. Bu bağlamda ilk akla gelen, oral sözcük dağarcığı alt-testidir. Bu çalışmada Düzey 1 testi, maddelerinde herhangi bir değişiklik yapılmaksızın (Amerikan para birimleri ile ilgili üç maddenin atılması dışında) Türkçe’ye çevrilmiş hali ile uygulanmıştır. Kültürden en fazla etkilenmeye aday olan oral sözcük dağarcığı alt-testi ile ilgili bulgular bu alt-testteki maddelerin yeniden ele alınması ve bazılarının değiştirilmesinin doğru olacağını göstermektedir. Örneğin, bu alt-testin içtutarlık açısından beklenenin altında bir düzeyde olduğu görülmüştür. Gelecekte bu alt-testin üzerinde çalışılarak iyileştirilmesi gerekmektedir.

Sonuç olarak, CogAT® Düzey 1 testinin Türkçe’ye çevrilmiş formunun geçerliğini saptamak ama-

cı ile yaptığımız bu çalışmanın bulguları, testin bu hali ile bile yüksek düzeyde geçerli olduğunu güçlü bir biçimde göstermektedir. Bundan sonra yapılması gereken, oral sözcük dağarcığı testinde yapılacak iyileştirmenin sonrasında, Düzey 1 testinin normlarının oluşturulmasıdır. Daha fazla sayıda ve ülkemizdeki ana sınıfı ve birinci sınıf öğrencilerini doğru olarak temsil etmek üzere seçilmiş okullarda, ders yılının değişik aralıklarında bu testin uygulandığı, daha etraflı bir çalışmanın yapılmasına gereksinim vardır. Normların saptanması testin öğretmenler ve okul yönetiminin, yukarıda dile getirilmiş ve diğer benzer amaçlar için, kullanılmasına olanak sağlayarak önemli bir eksiklik kapanmasına katkıda bulunacaktır. Ayrıca, normatif çalışmaların hem Birincil Batarya’nın Düzey 2 testi, hem de Çok-Düzeyle Batarya için yapılması, testin daha büyük yaşlardaki çocuklar için de kullanılmasına olanak sağlayacaktır. Geleneksel psikometrik yaklaşım bağlamında geliştirilmiş zeka testlerinde yararlanılmış olan zeka kurultusunu ülkemizde de ölçtüğü bu çalışmada gösterilmiş olan CogAT® grup testi, okul yaşındaki çocukların bilişsel yeteneklerinin ölçümünü gerektiren bilimsel araştırmalarda da kullanılabilir. Dolayısıyla bu test, eğitim sistemi içindeki yararlarının ötesinde, alanda kuramsal çalışmalar yapan diğer araştırmacılar için de yararlı olabilecek bir araçtır.

### Kaynaklar

- Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Upper Saddle River, IL: Prentice Hall.
- Demetriou, A. (2002). Tracing psychology’s invisible  $g_{\text{iam}}$  and its visible guards. In R. J. Sternberg & E. L. Grigorenko (Eds.). *The general factor of intelligence: How general is it?* (pp. 3-18). Mahwah, NJ: Lawrence Erlbaum.
- Embretson, S. E., & Schmidt McCollam, K. M. (2000). Psychometric approaches to understanding and measuring intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 423-444). Cambridge, UK: Cambridge University Press.
- Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.). *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53-91). New York: Guilford.

- Neisser, U., Boodoo, G., Bouchard (Jr.) ,T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Lochlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51(2), 77-101.
- Sternberg, R. J. (2002). Beyond g: The theory of successful intelligence. In R. J. Sternberg & E. L. Grigorenko (Eds.). *The general factor of intelligence: How general is it?* (pp. 447-479). Mahwah, NJ: Lawrence Erlbaum.
- Thorndike, R. L., & Hagen, E. P. (1993). *Form 5 CogAT®: Directions for administration Level 1*. Itasca, IL: Riverside.
- Thorndike, R. L., & Hagen, E. P. (1996). *Form 5 CogAT®: Interpretive guide for school administrators*. Itasca, IL: Riverside.



## *Summary*

# A Validation Study of CogAT® for Turkish Kindergarten and Grade 1 Students

I. Ercan Alp \*      Ayşesim Diri  
Boğaziçi Üniversitesi

Assessment of children's cognitive abilities is required in the educational system for a variety of reasons. For practical reasons group tests are required, and so far no group test of cognitive abilities have been standardized in Turkey. The present study was carried out as part of a larger project that aims at a standardization of Cognitive Abilities Test (CogAT®) in this culture. The objective of this phase of the project was an assessment of construct validity of the CogAT® (Form 5) Primary Battery for kindergarten and grade 1 students.

We hypothesized that test scores (1) differentiate between grade levels, (2) are affected by the quality of education provided by the school, and (3) predict current and later achievement in school. A fourth hypothesis was that the test has high internal consistency. All of these hypotheses are consistent with the intelligence construct as defined within the traditional psychometric approach and employed in the construction of the test. Empirical support for them would suggest that the test has high construct validity in this culture.

### **Method**

#### **Sample**

A total of 405 children (45 kindergarten and 360 grade 1 students) participated in the study. They were sampled from five, three private and two public, schools. Students in two of the private schools (School A and School B) were almost exclusively children of high SES families, and they

were instructed by highly qualified teachers in rather small classes. In both schools, a research unit effectively monitored the quality of teaching, and students' progress was evaluated by administering achievement tests on a regular basis. Results of these assessments were communicated both to teachers and parents. The third private school (School C) showed somewhat more variability in children's family background and we had the impression that the quality of education was not as high as in the other two. The remaining two schools (Schools D and E) were chosen to represent ordinary public schools in which academic standards generally are considerably lower. The only school in which we had access to kindergarten students was School B. We recorded children's gender in Schools A and B. Because the administration was carried out on four consecutive days late in the school year, the sample size was seriously reduced for the analyses performed on the composite score due to attendance problems. Data on the entire test was available only for 303, 27 kindergarten and 276 grade1, students. The distribution of grade 1 students for whom composite scores were available, across the five schools from School A to E, was 17, 38, 194, 11, and 16, respectively. Among those students for whom composite scores were available, the number of girls and boys for whom we also had gender information was 9 and 16, at kindergarten level, and 26 and 29 at grade 1 level, respectively.

#### **Measures**

CogAT® (Form 5) Primary Battery (Level 1) was

\*Address for Correspondence: I. Ercan Alp, Boğaziçi Üniversitesi, Psikoloji Bölümü, 34342 Bebek-İstanbul, Turkey.  
E-mail: alpe@boun.edu.tr

administered to all participants in Turkish. The Turkish form was different from the original in only one respect: three items, all involving American coins, were deleted from the quantitative concepts test. Thus, the total number of items in the test was 117 rather than 120. In addition to test scores, we also obtained some measure of concurrent achievement in Schools A and B. In the former, it was the grades given by teachers. In the latter, we had access to the ratings of the class teacher for kindergarten students (on a scale of 1 to 5) and scores on an achievement test administered by the school administration for grade 1 students. Three years later we further obtained course grade averages for that year for students who were still attending the same school in the case of all three private schools (Schools A, B, and C).

### **Procedure**

The authors translated the test into Turkish and trained the testers, all of whom were Psychology seniors at Bogazici University. In each of the participating schools, the test was administered on four consecutive days, beginning with the Practice Test on the first day and continuing with the verbal, quantitative, and nonverbal batteries on the following three days. Each administration was completed within one class hour, and instructions for administration were strictly adhered to. In each session, two testers were present. While one tester was carrying out the administration, the other moved around the room to check whether the instructions were followed by the students and answered any questions related to procedure or the mechanics of test-taking.

### **Results**

In the analyses reported here only raw scores were used. The mean item difficulty of the test, expressed in terms of percentage of correct answers, was 56.2 for kindergarten and 68.3 for grade 1 students. For each of the six subtests, we observed a reasonably wide range of item difficulty values and no item was answered correctly by all children at

either grade level. Gender information was obtained from the students enrolled in two schools (Schools A and B), and a comparison of the composite score of boys with that of girls showed no difference between them,  $t(78) = .94$ ,  $p = .35$ . Further comparisons, separately for kindergarten and grade 1 students, again failed to show a gender difference in composite scores,  $t(23) = .28$ ,  $p = .78$ , and  $t(53) = .81$ ,  $p = .42$ , respectively.

To assess the internal consistency of the test, KR20 coefficients were calculated at all three ability levels. At the highest level of ability, involving the entire test, and in the case of two batteries at the intermediate level, quantitative, and nonverbal, KD20 value was .97. The value found for the verbal battery was lower, .79. At the subtest level, the lowest value was found for the oral vocabulary test, .55, and the highest for the matrices test, .91. In general, these results demonstrate a high level of internal consistency for the test and provide support for the corresponding hypothesis.

In line with the second hypothesis, we predicted that private school students would obtain higher scores than public school students. Comparisons of grade 1 students from the two kinds of schools confirmed this prediction at all three levels of ability: for the composite scores,  $t(274) = 5.91$ ,  $p < .001$ ; verbal, quantitative, and nonverbal scores,  $t(358) = 9.53$ ,  $p < .001$ ,  $t(48.63) = 6.44$ ,  $p < .001$ , and  $t(71.92) = 8.41$ ,  $p < .001$ , respectively; and the six subtest scores, all  $t$ 's 5.63 or higher and all  $p$ 's  $< .001$  (whenever appropriate, a t-test that does not assume homogeneity of variances was performed and corresponding values are reported here).

Comparisons of individual schools within each group indicated that there were differences among them. In the case of private schools, School B students' scores at all three levels of ability were higher than those of School C students, and School A students' scores were not significantly different from those of either of the other two schools. The only exception to this pattern involved the relational concepts subtest, on which no difference was found



among the private schools. In the case of public schools, scores of School E students were consistently higher than those of School D students at all three ability levels without exception. Because School C was judged by us to be academically less competitive than Schools A and B, the results showing a difference between School B and C in favor of the former are consistent with our hypothesis that test scores are related with the quality of education provided by the school.

A third hypothesis predicted that grade 1 students obtain higher scores than kindergarten students. Because we had access to kindergarten students only in School B, we compared their scores only with grade 1 students attending the same school. Comparisons on all three levels of ability yielded significant differences in the predicted direction: composite scores,  $t(63) = 8.61, p < .001$ ; verbal, quantitative and nonverbal scores,  $t(90) = 6.63, p < .001$ ,  $t(76) = 5.75, p < .001$ , and  $t(73) = 9.41, p < .001$ , respectively; and individual subtest scores, all  $t$ 's  $> 3.65$ , all  $p$ 's  $< .001$ . These results confirmed our hypothesis that the test differentiates between age groups providing further support for the construct validity of the test.

That test scores predict current and later achievement was a fourth hypothesis tested in this study. In School A, the composite scores predicted course grade averages,  $r(16) = .63; p < .02$ . In School B, the composite scores of kindergarten students predicted class teachers' ratings,  $r(27) = .66, p < .001$ , and grade 1 students' composite scores predicted their average scores on the achievement test,  $r(37) = .68; p < .001$ . These results confirmed our hypothesis by providing evidence for the concurrent validity of the test.

In the case of all three private schools, we also had access to course grade averages of students who were still attending the same school three years after the CogAT® data were collected. Composite scores predicted later achievement of the students who were tested as grade 1 students,  $r(177) = .50, p <$

$.001$ , as well as those who were tested while in kindergarten,  $r(14) = .65, p < .02$ . When calculated separately for the individual schools, each of the correlation coefficients was significant at  $p < .001$  level,  $r(9) = .95$ ,  $r(30) = .61$ , and  $r(138) = .50$ , for Schools A to C, respectively. Thus, the results of this study also provide evidence for the predictive validity of the test for Turkish pupils.

### Discussion

Results of the present study showed that the translated form of CogAT® has a high level of internal consistency, and the scores (1) increase with age and quality of education provided in the school and (2) strongly predict current and later school achievement. Thus, all four hypotheses of the study were confirmed.

One of the weaknesses of the study was that we did not employ a quantitative measure of the quality of education variable. Although we took into account several aspects of quality to guide our selection of the schools (e.g., class size, existence of a research unit in the school to monitor students' progress, and how effectively that unit seems to work), our decisions were made on the basis of qualitative judgment. A formal measurement of the variable would lead to a confirmation of the hypothesized relation between test scores and quality of education beyond any doubt and help explaining more cogently the differences observed in this study between the individual schools within the public and private groups.

That we did not obtain data on the family background of individual children may be considered as a second weakness of the study. In this study, the quality of education provided in the school was confounded by the child's SES background. Although the majority of children in the private schools came from high SES families, there were some from lower levels who received free education in the school. Similarly, it is conceivable that some of the children in the public schools came from middle SES families. However, the conclusion we

drew about the relationship between test scores and quality of education received in school may be justified on the basis of three considerations: (1) we tried to select schools to represent a wide enough range of quality of education they provide, (2) CogAT® is a test specifically constructed to measure abilities that have been repeatedly shown to be affected by the quality of education, and (3) the test was administered to children at the end of the school year when the effect of the school should be already strong even for kindergarten pupils.

A third weakness of the study was that we had access to kindergarten pupils in one school only. This not only resulted in a very small sample size for younger students, but also prevented us from comparing different schools. Thus, it is not clear how general our findings are with respect to that level. However, the results of analyses on their data were fully consistent with those from grade 1 students and they all confirmed our hypotheses. Although independent replication of our findings are in order, we do not seriously doubt their generalizability.

Another problem was related to the timing of data collection. We witnessed a high level of absenteeism in the final weeks of school which led to a seriously reduced sample size for the analyses that involved the composite score. The number of children available for all the sessions dropped down to 303 from an initial group of 405 strong. However, results

based on scores from the individual batteries (verbal, quantitative, and nonverbal) each administered on a separate session, were consistent with those based on the entire test. Even though it might not have seriously affected the results of this study, we advise interested researchers not leave data collection late in the school year.

Despite the above mentioned weaknesses of this study, the results about the test's criterion validity, and especially predictive validity, are striking. It is worth mentioning that the approximate value Neisser et al. (1996) give for the correlation between intelligence test scores and grades, .50, is the same as found in this study for grade 1 students, despite a three-year interval between the two measurements.

Although we found strong evidence for the construct validity of the test in its present form, the results suggest that the quality of the test may be increased to an even higher level by improving the oral vocabulary subtest. Once that is done, a normative study of the Level 1 test of CogAT® Primary Battery is in order.

To conclude, the present study has introduced a group ability test to this culture that teachers and school administrators may find useful in their work at kindergarten and grade 1 levels. Moreover, the test may also aid researchers interested in the measurement of cognitive abilities in this age range.