

Bilgi İşlemsel Düşünme Becerisinin Deđerlendirilmesine İlişkin Sistemik Alanyazın Taraması

Ezgi TOSİK-GÜN 

Gazi Üniversitesi, Gazi Eğitim Fakóltesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü, Ankara, Türkiye.

ezgi.tosik@gmail.com

Tolga GÜYER 

Gazi Üniversitesi, Gazi Eğitim Fakóltesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü, Ankara, Türkiye.

tguyer@gmail.com

ARAŞTIRMA MAKALESİ/RESEARCH ARTICLE

Makale Bilgileri	ÖZ
<p>Makale Geçmişi Geliş: 27.07.2019 Kabul: 20.10.2019 Yayın: 20.12.2019</p>	<p>Yeni nesil bireylerde bulunması beklenen beceriler listesine teknolojik gelişmelere paralel olarak bilgi işlemsel düşünme (BİD) becerisi de eklenmiştir. Kısaca “teknolojiyi kullanarak problem çözmeye” olarak tanımlayabileceğimiz bu beceriyi geliştirmek için hazırlanan eğitimlerin etkilerinin belirlenmesi amacıyla sıra deđerlendirmeye geldiğinde, henüz kabul edilen, geçerli ve güvenilir yöntemlerin oluşmadığı görülmektedir. Bu nedenle alanyazındaki farklı deđerlendirme yöntemlerinin belirlenmesinin, farklılıkların ortaya konmasının ve bu yöntemlerin pozitif/negatif yönlerinin tartışılmasının geliştirilecek deđerlendirme yöntemleri için önemli katkı sağlayacağı düşünülmektedir. Bu amaçla Scholar Google, Web of Science ve ERIC veri tabanlarından dahil etme ve çıkarma kriterlerine uyan 47 araştırma analiz edilmek üzere seçilmiştir. Bu araştırmaların BİD becerisini deđerlendirmek amacıyla kullandığı yöntemler, yapılan içerik analizi sonucu şu başlıklarda incelenmiştir: i) en çok deđerlendirilen BİD becerisi bileşenleri, ii) veri toplama yöntemleri, iii) veri analiz yöntemleri, iv) veri toplama araçlarının içerikleri, v) hedef kitle ve vi) deđerlendirme yöntemlerine ait geçerlik/güvenirlik çalışmaları. Analiz sonuçlarına göre BİD becerisinin en fazla deđerlendirilen bileşenleri; soyutlama, algoritmik düşünme, ayırıştırma, test etme, hata ayıklama ve veri okuryazarlığı olarak sıralanmaktadır. Veri toplama yöntemi olarak ilk beş sırada görev, çoktan seçmeli soru, proje, açık uçlu soru ve görüşmenin kullanıldığı, toplanan verilerin çoğunlukla likert/rubriklerle analiz edildiği ortaya konmuştur. Veri toplama araçlarının içeriklerinin ise programlama, algı/tutum, matematik, günlük hayat problemleri ve genel yetenekten oluştuğu belirlenmiştir. Deđerlendirme yöntemlerinin hedef kitlesini en fazla K-12 seviyesi oluşturmuştur. Ayrıca incelenen 47 araştırmadan sadece altısına ait hem geçerlik, hem güvenilirlik çalışmalarına ulaşılmıştır. Belirlenen birbirinden farklı deđerlendirme yöntemleri karşılaştırılmış, avantajlı ve dezavantajlı yönleri tartışılmıştır. Elde edilen sonuçlara göre BİD becerisini deđerlendirme yöntemlerine yönelik gelecek araştırmalar için öneriler sunulmuştur.</p>
<p>Anahtar Kelimeler: Bilgi İşlemsel Düşünme Becerisi, Deđerlendirme, Sistemik Alanyazın Taraması.</p>	

A Systematic Literature Review on Assessing Computational Thinking

Article Info	ABSTRACT
<p>Article History <i>Received: 27.07.2019</i> <i>Accepted: 20.10.2019</i> <i>Published: 20.12.2019</i></p> <p>Keywords: Computational Thinking, Assessment, Systematic Literature Review.</p>	<p>Computational thinking (CT) has been added to the list of expected skills from individuals in parallel with technological advances. There are no fully-accepted, valid, and reliable methods to evaluate this skill which can be defined as "problem solving with using technology". Therefore, identifying the different assessment methods, revealing their differences, and discussing the positive/negative aspects of these methods will contribute to the new assessment methods to be developed. In order to achieve this purpose, 47 studies that meet the inclusion/exclusion criteria from Scholar Google, Web of Science and ERIC databases were selected for analysis. After the content analysis, these studies were examined under the following headings: i) the most evaluated CT components, ii) data collection methods, iii) data analysis methods, iv) content of data collection tools, v) audience, and vi) validity and reliability studies. According to the results of the analysis, the most evaluated components of CT can be listed as; abstraction, algorithmic thinking, decomposition, testing, debugging and data literacy. The top five data collection methods were tasks, multiple choice questions, projects, open-ended questions, and interviews. The data collected by using these methods were mostly analyzed with Likert scale/rubrics. The contents of the data collection tools consist of programming, perception/attitude, mathematics, daily life problems, and general ability. The target group of the assessment methods was mostly the K-12 level. Additionally, there are 6 out of 47 studies which include both validity and reliability were obtained. The advantages/disadvantages of different assessment methods were discussed. Also, suggestions for future studies about the methods of evaluating CT are presented.</p>

GİRİŞ

Teknolojik gelişmelerin hız kazandığı ve makineleşmenin arttığı günümüzde teknolojiyi yalnızca tüketen bireylerden oluşan toplumların global yarışta etkin yer alması düşünülemez. Her dönemde bu yarışa dahil olmak isteyen toplumların eğitim politikaları, ihtiyaç duyulan insan profilini yetiştirmeyi amaçlamaktadır. Bu doğrultuda öncelikle bireylerin sahip olması gereken bilgi ve becerilerin belirlenmesi gerekmektedir. Son dönemde sıklıkla 21. yy. becerileri kapsamında duyduğumuz yaratıcılık, iletişim, iş birliği, girişimcilik, problem çözme... vb yeni nesil bireylerde bulunması gereken beceriler olarak listelenmektedir (OECD, 2018). Teknolojik gelişmelere paralel olarak, bu listeye kısaca "teknolojiyi kullanarak problem çözme" olarak tanımlayabileceğimiz bilgi işlemsel düşünme (BİD) becerisi de eklenmiştir (Gretter ve Yadav, 2016).

Yine son dönemde oldukça üzerinde durulan ve giderek yaygınlaşan fen, teknoloji, mühendislik ve matematik disiplinlerinin bütünleşik bir programda sunulduğu eğitimlerde ve kendin yap (do it yourself-DIY) projelerinin teknoloji ile birleştirildiği maker hareketi kapsamında bireylerin teknolojiyle üretim becerilerinin geliştirilmesi hedeflenmektedir (Bakırcı ve Kutlu, 2018). Birçok farklı becerinin işe koşulduğu bu süreçte BİD becerisinin yeri oldukça önemlidir (Grover ve Pea, 2013). Wing (2006)'in açıklamalarıyla popülerlik kazanan bu kavram aslında yeni olmayan bir tür problem çözme sürecidir ve Yağcı (2019) ve Korkmaz, Çakır ve Özden'e (2017) göre bu beceri algoritmik düşünme, yaratıcılık ve eleştirel düşünme becerilerini gibi farklı becerileri kapsamaktadır. Günlük hayatın her aşamasında problem çözme var olduğuna göre, yaşanan çağın teknolojik altyapısını kullanarak çözümler üretmek, belirli bir meslek grubundan değil, yeni nesil her bireyden beklenmektedir (Wing, 2006; ISTE, 2014). Bu beklentiyi karşılamak amacıyla araştırmacılar tarafından BİD becerisini geliştirmesi öngörülen, bilgisayar bilimleri ve

programlama içeriklerinin yoğun olarak kullanıldığı farklı disiplinlerde etkinlikler oluşturulmuştur. Örneğin; Bell, Alexander, Freeman ve Grimley (2009) ve Caldwell ve Smith (2016) bilgisayarsız bilgisayar bilimleri, Kim, Kim ve Kim (2013) bilgisayarsız programlama, Bers, Flannery, Kazakoff ve Sullivan (2014) robotik ve programlama, Sengupta, Kinnebrew, Basu, Biswas ve Clark (2013) programlama, fizik ve biyoloji temelli etkinlikler geliştirmişlerdir.

Hazırlanan eğitim içeriklerinin BİD becerisi üzerindeki etkisinin belirlenmesi amacıyla sıra değerlendirmeye geldiğinde henüz kabul edilen geçerli ve güvenilir yöntemlerin oluşmaması gibi bir sorunla karşılaşmaktadır (Shute, Sun ve Asbell-Clarke, 2017; Lockwood ve Mooney, 2018). Bu durumun nedenlerini, kavramın çok yeni olması, kuramsal temelleri hakkındaki çalışmaların dahi henüz devam ediyor olması ve karmaşık üst düzey bir beceri olması olarak sıralamak mümkündür (Weinberg, 2013). Oysa eğitimde ölçme ve değerlendirmenin belirlenen hedefe ne derece ulaşıldığının tespit etme, eksikleri belirleme, geri bildirim verme ve düzenlemeler yapma gibi birçok açıdan önemi büyüktür (Katchapakirin ve Anutariya, 2018). BİD becerisinin değerlendirilmesinde yaşanan problemler bu becerinin nasıl geliştirilebileceği ile ilgili objektif kararların alınmasının önünde engel oluşturmaktadır. Kabul edilen, geçerli ve güvenilir değerlendirme yöntemlerinin oluşmaması nedeniyle alanyazındaki mevcut çeşitliliğin belirlenmesinin ve var olan yöntemlerin pozitif ve negatif yönlerinin tartışılmasının geliştirilecek değerlendirme yöntemleri için önemli katkı sağlayacağı düşünülmektedir.

Bu doğrultuda, bu araştırmanın amacı: BİD becerisinin değerlendirilmesine yönelik alanyazında yer alan farklı yöntemlerinin belirlenmesi, farklılıkların ortaya konulması ve bu yöntemlerin pozitif-negatif yönlerinin tartışılmasıdır. Bu amaç doğrultusunda aşağıdaki araştırma sorusuna cevap aranacaktır:

S. Alanyazında BİD becerisinin değerlendirilmesine yönelik kullanılan farklı yöntemler nelerdir?

İLGİLİ ALANYAZIN VE ÖZGÜN DEĞER

BİD becerisini farklı değişkenler altında inceleyen sistematik derleme çalışmaları yer almaktadır. Hsu, Chang ve Hung (2018) BİD becerisinin öğretilmesi üzerine yapılan araştırmaları konu almışlardır. 2006 ve 2017 yılları arasında, belirlenen kriterlere uyan 120 araştırma seçilmiştir. Bu araştırmaların analizlerinde kullanılan başlıklar: uygulama alanı, öğrenme stratejileri, katılımcılar, programlama dili, ders kategorisi ve öğretme aracı olarak sıralanmıştır. Kalelioğlu, Gülbahar ve Kukul (2016) ise seçtikleri 125 araştırmayı çalışmanın amacı, hedef kitlesi, teorik temelleri, önerdiği BİD tanımı, kapsamı, araştırma türü ve deseni olmak üzere 6 kategoride incelemişlerdir. Bu araştırmalarda BİD becerisinin değerlendirilmesine ilişkin herhangi bir analiz yer almamaktadır.

BİD becerisinin değerlendirilmesine yönelik analizlerin yer aldığı sistematik derleme çalışmaları araştırıldığında, yaş grupları ya da konu alanı gibi belirli değişkenler açısından inceleme yapan sınırlı sayıda araştırma ile karşılaşmıştır. Lockwood ve Mooney'in (2018) 59 araştırma ile yürüttükleri çalışma incelendiğinde ortaöğretimde BİD becerisinin geliştirilmesi amacıyla hangi konu alanlarının kullanıldığı, hangi araç ve yöntemlerin geliştirildiği ve hangi yöntemlerle değerlendirildiği dikkate alınarak analizler yapıldığı görülmektedir. Değerlendirme ile ilgili başlığın sonuçları, belirlenen beş farklı yöntem hakkında özet bilgiler içermektedir. Bizim araştırmamıza benzer olarak De Araujo, Andrade ve Guerrero (2016) da BİD becerisinin değerlendirilmesine yönelik alanyazında yer alan yöntemleri incelemişlerdir. 8 makale 19 konferans sunumundan oluşan 27 çalışma üç başlık altında analiz edilmiştir. Bunlar: i) BİD

becerisini geliştirmek için hangi pedagojik yaklaşım, hangi hedef kitle için seçilmiştir? ii) BİD becerisi kapsamında hangi yetenekler ölçülmüştür? iii) BİD becerisini ölçmek amacıyla kullanılan araçlar nelerdir? Elde edilen sonuçlara göre i) en çok kullanılan pedagojik yaklaşımın belirli bir süreci kapsayan eğitimler olduğu ve en fazla K-12 öğrencileri ile çalışıldığı, ii) BİD becerisini ölçmek amacıyla en fazla problem çözme, algoritma ve soyutlama becerilerinin değerlendirildiği ve iii) en fazla kullanılan araçların kodlama ve çoktan seçmeli testler olduğu belirlenmiştir. Sondakh (2018) ise BİD becerisi ile ilgili yükseköğretim seviyesindeki 41 araştırmayı incelemiştir. Bu çalışmalardan %20'sinin değerlendirmeyi, %60'ının eğitim içeriklerini ve %20'sinin de her ikisini de konu aldığını ifade etmiştir. Sonuçlar en fazla değerlendirilen BİD becerisi bileşenlerinin algoritmalar ve soyutlama olduğunu ortaya koymuştur.

Alanyazında yer alan sistemik derleme araştırmaları göz önüne alındığında, çalışmaların genellikle BİD becerisi kavramını ve kapsamını ortaya koymaya yönelik olduğu görülmektedir. BİD becerisinin değerlendirilmesi ile ilgili ulaşılan kaynaklar ortaöğretimle ve yükseköğretimle sınırlandırılmış veya en son 2016 yılına ait verileri içermektedir. Bütün eğitim seviyelerini kapsayan, araştırmaların farklı başlıklarda sentezlendiği, güncel verilerin dahil edildiği ve farklı veri tabanlarında yapılacak sistemik derleme çalışmalarının alanyazına katkı sağlayacağı düşünülmektedir.

YÖNTEM

Alanyazındaki birbirinden farklı değerlendirme yöntemlerinin ortaya konmasını ve bu yöntemlerin BİD becerisini ölçme konusunda yeterliliklerinin tartışılmasını amaçlayan bu araştırmada sistemik alanyazın taraması yapılmıştır. Sistemik alanyazın taraması belirlenen araştırma sorusunu cevaplamak amacıyla yapılmış araştırmaların kapsamlı olarak taranması, belirli kriterlere göre elemelerin yapılması ve seçilen çalışmalardan sentez elde edilme sürecidir (Kitchenham, 2004; Andrews ve Harlen, 2006). Sistemik alanyazın taramasında izlenen bütün basamakların net olarak açıklanması, bu yöntemin yanlılıktan kurtulan, tekrar edilebilir ve doğrulanabilir bir yöntem olarak kabul görmesini sağlamaktadır (Karaçam, 2013).

Tarama Süreci ve İncelenen Araştırmalar

İncelenecek çalışmalar belirlenirken Web of Science (WOS), ERIC ve Scholar Google veritabanlarında tarama yapılmıştır. Yalnızca BİD becerisinin değerlendirilmesine yönelik araştırmaların sayısı oldukça azdır. Genellikle BİD becerisini geliştirmeye yönelik yapılan çalışmalarda değerlendirme yöntemlerine yer verildiği ve bu araştırmaların başlıklarında değerlendirmeye yönelik bir terim bulunmadığı görülmüştür. Ayrıca İngilizcede, Türkçede olduğu gibi BİD becerisini karşılayan farklı terimler kullanılmamaktadır. Bu nedenlerle anahtar kelime olarak yalnızca “computational thinking” seçilmiş, “assessment” gibi ilişkili kelimeler aramaya dahil edilmemiştir. Tarama, çalışmaların başlıklarında yapılmış ve 2018 yılıyla sınırlandırılmıştır. WOS'ta 518, ERIC'te 106 ve Scholar Google'da 997 araştırma dâhil etme ölçütlerine göre taranmıştır.

Dâhil Etme ve Çıkarma Ölçütleri

Hangi araştırmaların dâhil edileceğine ve hangi araştırmaların kapsam dışına çıkarılacağına karar vermek amacıyla aşağıdaki ölçütler göz önüne alınmıştır.

Dâhil etme ölçütleri:

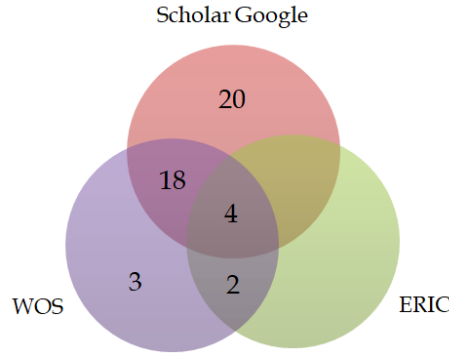
1. Dili İngilizce olan,

2. Tam metinlerine erişilebilen,
3. Özetin okunmasıyla BİD becerisinin değerlendirilmesini kapsadığı düşünülen 157 araştırma detaylı okuma için seçilmiştir.

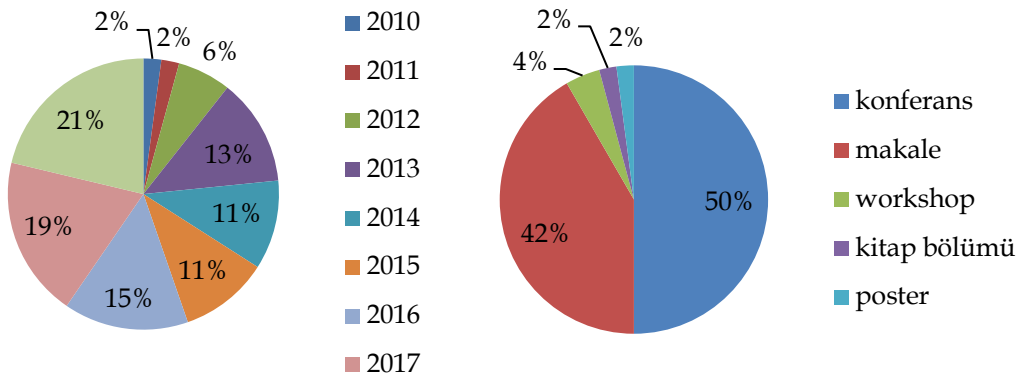
Çıkarma ölçütleri:

4. Veri toplama yöntemi, ölçme aracının içeriği ve kavramsal çerçeve başlıklarına ait açıklamaları eksik olan (27),
 5. Önceden belirlenmiş bir yöntemi tekrar kullanan (38),
 6. Kavramsal çerçevesinde yer alan bileşenlerden herhangi birini değerlendirme dışı bırakan (15),
 7. BİD becerisini “ayrıştırma nedir, BİD becerisi nedir?” şeklinde kavramsal olarak değerlendiren (7),
 8. BİD becerisinin değerlendirilmesiyle ilgisi olmadığı belirlenen (23)
- toplam 110 çalışma kapsam dışı bırakılmıştır.

Detaylı okuma sürecinde çıkarma ölçütlerinde açıklanan araştırmalar ihmal edilmiş ve 47 araştırma analiz edilmiştir. Seçilen araştırmalar ve hangi veritabanlarında yer aldıklarına dair bilgiler Şekil 1’de yer almaktadır. Şekil 2’de ise, analiz edilen araştırmaların yıllara ve türlerine göre dağılımları gösterilmektedir.



Şekil 1. İncelenen araştırmaların sayısı ve tarandıkları veritabanları



Şekil 2. İncelenen araştırmaların a) yıllara göre dağılımı b) türlerine göre dağılımları

Şekil 2’de en fazla 2018 yılında yayınlanan araştırmaların incelendiği, incelenen araştırmaların ise, yarısının konferans metinlerinden oluştuğu görülmektedir.

Veri Analizi, Geçerlik ve Güvenirlik

Bir araştırmacı tarafından 47 çalışmaya içerik analizi yapılmıştır. Her çalışma veri toplama yöntemi, ölçme aracının içeriği, kavramsal çerçeve, hedef kitle ve geçerlik/güvenirlik başlıklarında detaylı olarak incelenmiştir. Bu süreçte belirtilen başlıklara ve çalışmaların farklılaştığı noktalara ait notlar alınmıştır. Alınan notlar birbiriyle karşılaştırılmış, benzer olanlar temalaştırılmış, farklılaşan durumlar için yeni temalar oluşturulmuştur (Elo ve Kyngäs, 2008). İç güvenirliliği artırmak için tutarlılık incelenmesi yapılmıştır (Guba ve Lincoln, 1982). Bu amaçla temaların bütün araştırmalarla tutarlılık gösterdiğinden emin olmak ve tek kodlayıcı güvenirliliğini (intra-rater reliability) sağlamak için iki ay sonra 47 çalışma aynı araştırmacı tarafından yeniden okunmuştur (Jonsson ve Svingby, 2007; Ergai vd., 2016). Yapılan iki kodlama arasındaki uyum Cohen Kappa katsayısıyla test edilmiştir (McHugh, 2012). Elde edilen 0,82 değeri, Landis ve Koch'a (1977) göre 0,81 ve 1,00 aralığında çok iyi düzeyde uyumun olduğu göstermektedir. Dış güvenirliliği artırmak amacıyla teyit edilebilirlik sağlanmaya çalışılmıştır (Streubert ve Carpenter, 2011). Bu doğrultuda bulgulara ait tablolar, araştırmaların künye bilgilerini de içermektedir. Ayrıca araştırmaya ait bütün süreçlerin ayrıntılı betimlesinin yapılarak dış geçerliğin artırılması hedeflenmiştir (Guba ve Lincoln, 1982).

BULGULAR

S. Alanyazında BİD becerisinin değerlendirilmesine yönelik kullanılan farklı yöntemler nelerdir?

Yapılan içerik analizi sonucu incelenen değerlendirme yöntemlerinin kavramsal çerçeve, veri toplama yöntemleri, veri analiz yöntemleri, veri toplama araçlarının içerikleri, hedef kitle ve geçerlik/güvenirlik çalışmaları başlıklarında farklılaştığı tespit edilmiştir. Bu kısımda bahsedilen başlıklara ait detaylı bulgular sunulmuştur. Belirledikleri yöntemi tekrar kullanmak isteyen araştırmacıların kaynaklara ulaşmasını kolaylaştırmak adına bütün tablolar araştırma künyeleriyle birlikte verilmiştir.

BİD Becerisinin Değerlendirilmesinde Kullanılan Bileşenler

Alanyazında BİD becerisinin kavramsal çerçevesinin henüz netleşmemiş olmasından dolayı araştırmacıların kabul ettikleri bileşenler incelenmiştir. Çalışmaların BİD becerisi kapsamında değerlendirdikleri bileşenlere ait bulgular Tablo 1'de gösterilmektedir.

Tablo 1. BİD Becerisinin Değerlendirilmesinde Sıklıkla Kullanılan Bileşenler

BİD becerisi bileşenleri	f	Araştırma
Soyutlama	26	Brennan ve Resnick, 2012; Games ve Kane, 2012; Werner, Denner, Campe ve Kawamoto, 2012; Gouws vd., 2013a; Gouws vd., 2013b; Seiter ve Foreman, 2013; Atmatzido ve Demetriadis, 2014; Wang, Wang ve Liu, 2014; Moreno-León ve Robles, 2015; Nesiba, Pontelli ve Staley, 2015; Atmatzidou ve Demetriadis, 2016; Dagiene ve Stupuriene, 2016; Hoover vd., 2016; Leonard vd., 2016; Zhong, Wang, Chen ve Li, 2016; Grover vd., 2017; Rowe, Asbell-Clarke, Cunningham ve Gasca, 2017; Weese ve Feldhausen, 2017; Witherspoon vd., 2017; Zhang ve Biswas, 2017; Anistiyasari ve Kurniawan, 2018; Chang, Sun, Wu ve Guizani, 2018; Katchapakirin ve Anutariya, 2018; Leonard vd., 2018; Ling, Saibin, Naharu, Labadin ve Aziz, 2018; Looi, How, Longkai, Seow ve Liu, 2018
Algoritmik düşünme	26	Games ve Kane, 2012; Werner vd., 2012; Aiken vd., 2013; Gouws, Bradshaw ve Wentworth, 2013a; Gouws, Bradshaw ve Wentworth, 2013b; Philip, Renumol ve Gopeekrishnan, 2013; Seiter ve Foreman, 2013; Atmatzido ve Demetriadis, 2014; Lee, Mauriello, Ahn ve Bederson, 2014; Grover, 2015; Nesiba vd., 2015; Atmatzidou ve Demetriadis, 2016; Dagiene ve Stupuriene, 2016; Leonard vd., 2016; Chen vd., 2017; Grover vd., 2017; Korkmaz vd., 2017; Rowe vd., 2017; Weese ve Feldhausen,

		2017; Witherspoon, Higashi, Schunn, Baehr ve Shoop, 2017; Hutchins vd., 2018; Leonard vd., 2018; Ling vd., 2018; Looi vd., 2018; Pérez-Marín, Hijón-Neira, Bacelo ve Pizarro, 2018; Proctor ve Blikstein, 2018
Ayrıştırma	14	Games ve Kane, 2012; Seiter ve Foreman, 2013; Atmatzido ve Demetriadis, 2014; Lee vd., 2014; Wang vd., 2014; Moreno-León ve Robles, 2015; Atmatzidou ve Demetriadis, 2016; Hoover vd., 2016; Rowe vd., 2017; Weese ve Feldhausen, 2017; Zhang ve Biswas, 2017; Anistyasari ve Kurniawan, 2018; Chang vd., 2018; Looi vd., 2018
Test etme ve hata ayıklama	14	Brennan ve Resnick, 2012; Bers vd., 2014; Wang vd., 2014; Falloon, 2015; Basogain vd., 2016; Leonard vd., 2016; Zhong vd., 2016; Grover vd., 2017; Weese ve Feldhausen, 2017; Zhang ve Biswas, 2017; Hutchins vd., 2018; Katchapakirin ve Anutariya, 2018; Leonard vd., 2018; Looi vd., 2018
Veri okuryazarlığı	12	Brennan ve Resnick, 2012; Games ve Kane, 2012; Seiter ve Foreman, 2013; Moreno-León ve Robles, 2015; Basogain vd., 2016; Hoover vd., 2016; Orton vd., 2016; Zhong vd., 2016; Chen vd., 2017; Anistyasari ve Kurniawan, 2018; Chang vd., 2018; Katchapakirin ve Anutariya, 2018
Sıralama	11	Brennan ve Resnick, 2012; Seiter ve Foreman, 2013; Bers vd., 2014; Falloon, 2015; Nesiba vd., 2015; Román-González, 2015; Zhong vd., 2016; Grover vd., 2017; Zhang ve Biswas, 2017; Katchapakirin ve Anutariya, 2018; Pérez-Marín vd., 2018
Akış kontrol yapıları	11	Brennan ve Resnick, 2012; Seiter ve Foreman, 2013; Werner vd., 2014; Nesiba vd., 2015; Román-González, 2015; Zhong vd., 2016; Grover vd., 2017; Zhang ve Biswas, 2017; Hutchins vd., 2018; Katchapakirin ve Anutariya, 2018; Pérez-Marín vd., 2018
Mantıksal düşünme	10	Gouws vd., 2013a; Gouws vd., 2013b; Kim vd., 2013; Moreno-León ve Robles, 2015; Leonard vd., 2016; Hoover vd., 2016; Anistyasari ve Kurniawan, 2018; Chang vd., 2018; Leonard vd., 2018; Ling vd., 2018
Paralleleştirme	10	Brennan ve Resnick, 2012; Seiter ve Foreman, 2013; Moreno-León ve Robles, 2015; Hoover vd., 2016; Zhong vd., 2016; Grover vd., 2017; Weese ve Feldhausen, 2017; Anistyasari ve Kurniawan, 2018; Katchapakirin ve Anutariya, 2018; Chang vd., 2018
Otomasyon	9	Brennan ve Resnick, 2012; Atmatzido ve Demetriadis, 2014; Wang vd., 2014; Falloon, 2015; Atmatzidou ve Demetriadis, 2016; Zhong vd., 2016; Grover vd., 2017; Zhang ve Biswas, 2017; Katchapakirin ve Anutariya, 2018
Örüntü tanıma	8	Games ve Kane, 2012; Gouws vd., 2013a; Gouws vd., 2013b; Lee vd., 2014; Dagiene ve Stupuriene, 2016; Rowe vd., 2017; Ling vd., 2018; Pérez-Marín vd., 2018
Modelleme	8	Werner vd., 2012; Aiken vd., 2013; Gouws vd., 2013a; Gouws vd., 2013b; Orton vd., 2016; Grover vd., 2017; Zhang ve Biswas, 2017; Bati, Yetişir, Çalışkan, Güneş ve Gül Saçan, 2018
İşlevsel yapılar	8	Brennan ve Resnick, 2012; Werner vd., 2014; Román-González, 2015; Basogain vd., 2016; Zhong vd., 2016; Grover vd., 2017; Hutchins vd., 2018; Katchapakirin ve Anutariya, 2018; Proctor ve Blikstein, 2018
Genelleme	7	Games ve Kane, 2012; Atmatzido ve Demetriadis, 2014; Lee vd., 2014; Atmatzidou ve Demetriadis, 2016; Dagiene ve Stupuriene, 2016; Leonard vd., 2016; Looi vd., 2018
Değişkenler	7	Philip vd., 2013; Seiter ve Foreman, 2013; Werner vd., 2014; Nesiba vd., 2015; Román-González, 2015; Zhang ve Biswas, 2017; Hutchins vd., 2018
Akış kontrolü	6	Bers vd., 2014; Moreno-León ve Robles, 2015; Hoover vd., 2016; Weese ve Feldhausen, 2017; Anistyasari ve Kurniawan, 2018; Chang vd., 2018
Problem çözme becerisi	5	Yeh, Xie ve Ke, 2011; Leonard vd., 2016; Korkmaz vd., 2017; Rowe vd., 2017; Leonard vd., 2018
Senkronizasyon	5	Seiter ve Foreman, 2013; Moreno-León ve Robles, 2015; Hoover vd., 2016; Anistyasari ve Kurniawan, 2018; Chang vd., 2018

Tablo 1’de yer alan bileşenlerden bazıları, benzerlik gösteren bileşenlerin birleştirilmesinden elde edilmiştir. Akış kontrol yapılarında; koşullar, döngüler ve iterasyonlar, işlevsel yapılarda; operatörler, fonksiyonlar ve eventler, son olarak otomasyonda ise, yeniden kullanma ve modülerlik bileşenleri birleştirilmiştir.

BİD becerisinin değerlendirilmesinde en fazla kullanılan bileşenlerin soyutlama, algoritmik düşünme, ayrıştırma, test etme ve hata ayıklama, veri okuryazarlığı, sıralama, akış kontrol yapıları olduğu Tablo 1’de görülmektedir. Listenin tamamı incelendiğinde BİD becerisinin bileşenleri olarak görülen kavramlar için tablo 2’deki gibi bir sınıflamanın yapılabileceği farkedilmiştir.

Tablo 2. BİD Becerisine Ait Alanyazında Yer Alan Bileşenlerin Genel Sınıflaması

Sınıflama	BİD becerisi bileşenleri
BİD becerisine özgü kavramlar	Soyutlama, ayırıştırma, otomasyon, örüntü tanıma, paralelizasyon, genelleme, modelleme...
21. yy. becerileri	Mantıksal düşünme, problem çözme, yaratıcılık, algoritmik düşünme, veri okuryazarlığı...
Programlama	Döngüler, değişkenler, koşullar, operatörler, fonksiyonlar...

Tablo 2’de görülen sınıflamanın yanısıra incelenen çalışmalarda tekrar kullanılan kavramsal çerçevelerin olup olmadığı analiz edilmiştir. Araştırmaların beşinde Brennan ve Resnick’in (2012), dördünde ise Dr. scratch analiz aracının bileşenlerinin kavramsal çerçeve olarak kabul edildiği belirlenmiştir.

Veri Toplama Yöntemleri

Yapılan analiz sonucunda BİD becerisini değerlendirmek amacıyla hangi araştırmaların hangi veri toplama yöntemini kullandıklarına dair bulgular Tablo 3’te sunulmuştur.

Tablo 3. Kullanılan Veri Toplama Yöntemleri

Veri toplama yöntemi	Kullanılan teknik	f	Araştırma
Görev		17	Games ve Kane, 2012; Werner vd., 2012; Aiken vd., 2013; Gouws vd., 2013a; Bers vd., 2014; Wang vd., 2014; Falloon, 2015; Grover, 2015; Zhong vd., 2016; Grover vd., 2017; Romero, Lepage ve Lille, 2017; Zhang ve Biswas, 2017; Hutchins vd., 2018; Katchapakirin ve Anutariya, 2018; Leonard vd., 2018; Looi vd., 2018; Proctor ve Blikstein, 2018
Çoktan seçmeli soru		14	Gouws vd.,2013b; Kim vd., 2013; Grover, 2015; Nesiba vd., 2015; Román-González, 2015; Basogain vd., 2016; Dagiene ve Stupuriene, 2016; Orton vd., 2016; Chen vd., 2017; Witherspoon vd., 2017; Zhang ve Biswas, 2017; Bati vd., 2018; Hutchins vd., 2018; Pérez-Marín vd., 2018
Proje		13	Koh, Basawapatna, Bennett ve Repenning, 2010; Brennan ve Resnick, 2012; Seiter ve Foreman, 2013; Bers vd., 2014; Werner, Denner ve Campe, 2014; Grover, 2015; Moreno-León ve Robles, 2015; Basogain vd., 2016; Hoover vd., 2016; Leonard vd., 2016; Anistiyasari ve Kurniawan, 2018; Chang vd., 2018; Proctor ve Blikstein, 2018
Açık uçlu soru		13	Yeh vd., 2011; Aiken vd., 2013; Gouws vd., 2013b; Kim vd., 2013; Philip vd., 2013; Atmatzido ve Demetriadis, 2014; Grover, 2015; Atmatzidou ve Demetriadis, 2016; Orton vd., 2016; Zhong vd., 2016; Chen vd., 2017; Zhang ve Biswas, 2017; Hutchins vd., 2018
Görüşme	Geleneksel	3	Wang vd., 2014; Hoover vd., 2016; Looi vd., 2018
	Sesli düşünme	3	Aiken vd., 2013; Lee vd., 2014; Atmatzidou ve Demetriadis, 2016
	Ürün temelli	2	Brennan ve Resnick, 2012; Grover, 2015
	Tasarım senaryosu	1	Brennan ve Resnick, 2012
	Etkileşimli	1	Games ve Kane, 2012
Sistem kayıtları		6	Games ve Kane, 2012; Falloon, 2015; Grover vd., 2017; Rowe vd., 2017; Katchapakirin ve Anutariya, 2018; Looi vd., 2018

Gözlem	4	Games ve Kane, 2012; Wang vd., 2014; Grover vd., 2017; Looi vd., 2018
Anket	4	Jun, Han ve Kim, 2017; Korkmaz vd., 2017; Weese ve Feldhausen, 2017; Leonard vd., 2018

Tablo 3'te en fazla kullanılan veri toplama yöntemlerinin sırasıyla görev, çoktan seçmeli soru, proje, açık uçlu soru ve görüşme olduğu görülmektedir. İzleme verilerini (sistemde dolanma süresi, etkileşim sayısı vb.) ve videoları (öğrencinin veya ekranının kaydı) kapsayan sistem kvayıtları, gözlem ve likert tipi anketler ise daha az kullanılmaktadır.

Veri toplama yöntemleri içerisindeki bazı ölçme araçlarının farklı araştırmacılar tarafından da kullanıldığı görülmüştür. Bu araçlar çoktan seçmeli sorular kategorisinde Román-González'in (2015) CTt ve uluslararası Bebras görevleri, likert kategorisinde ise Korkmaz ve arkadaşları (2017) tarafından geliştirilmiş "bilgisayarca düşünme beceri düzeyleri ölçeği" dir.

BİD becerisini değerlendirmek için 22 araştırmada birden fazla, 24 araştırmada ise tek veri toplama yöntemi tercih edilmiştir. Geriye kalan bir araştırma farklı yöntemlerle toplanan verilerin analiz edilmesi için yapılmış rubrik geliştirme çalışmasıdır.

Veri Analiz Yöntemleri

İncelenen araştırmalarda BİD becerisini değerlendirmek amacıyla toplanan verileri analiz etmek için kullanılan yöntemlerin farklılaştığı ve BİD becerisine özgü analiz yöntemlerinin oluşturulduğu belirlenmiştir. Araştırmacıların veri toplama yöntemlerine göre kullandıkları analizler Tablo 4'te yer almaktadır.

Tablo 4. Veri Toplama Yöntemlerine Göre Kullanılan Analiz Yöntemleri

Veri toplama yöntemi	Veri analiz yöntemi	f	Araştırma
Görev	Likert/Rubrik	11	Werner vd., 2012; Gouws vd., 2013a; Bers vd., 2014; Grover, 2015; Zhong vd., 2016; Grover vd., 2017; Zhang ve Biswas, 2017; Hutchins vd., 2018; Leonard vd., 2018; Looi vd., 2018; Proctor ve Blikstein, 2018
	Yazılım araçları	2	Romero vd., 2017; Katchapakirin ve Anutariya, 2018
	Sınıflama-Yüzde	1	Aiken vd., 2013
Proje	Yazılım araçları	6	Koh vd., 2010; Brennan ve Resnick, 2012; Moreno-León ve Robles, 2015; Hoover vd., 2016; Anistyasari ve Kurniawan, 2018; Chang vd., 2018
	Likert/Rubrik	6	Seiter ve Foreman, 2013; Bers vd., 2014; Basogain vd., 2016; Leonard vd., 2016; Romero vd., 2017; Proctor ve Blikstein, 2018
	Kontrol listesi	1	Werner vd., 2014
Açık uçlu sorular	Doğru-Yanlış	6	Yeh vd., 2011; Gouws vd., 2013b; Kim vd., 2013; Philip vd., 2013; Orton vd., 2016; Zhang ve Biswas, 2017
	Likert/Rubrik	5	Atmatzido ve Demetriadis, 2014; Atmatzidou ve Demetriadis, 2016; Zhong vd., 2016; Chen vd., 2017; Zhang ve Biswas, 2017
	Sınıflama-yüzde	1	Aiken vd., 2013
Görüşme	Likert/Rubrik	2	Atmatzidou ve Demetriadis, 2016; Hoover vd., 2016
	Tümevarımsal gömülü teori	1	Lee vd., 2014
	Söylem analizi	1	Games ve Kane, 2012

Sistem kayıtları	Kontrol listesi	1	Rowe vd., 2017
	Yazılım araçları	1	Falloon, 2015
	Söylem analizi	1	Games ve Kane, 2012
Gözlem	Likert/Rubrik	1	Grover vd., 2017
	Söylem analizi	1	Games ve Kane, 2012
	Rubrik	1	Ling vd., 2018

Tablo 4'te görüldüğü gibi, farklı yöntemlerle toplanan verilerin analizinde likert veya rubriklerle puanlama yöntemi sıklıkla kullanılmaktadır. Games ve Kane (2012) gözlem, görüşme ve sistem kayıtlarından elde ettikleri verileri bütüncül olarak söylem analizi ile incelemişlerdir. Ling ve arkadaşları (2018), farklı veri türlerinin puanlanması için BİD becerisine özgü rubrik geliştirme çalışması yapmışlardır.

Likert veya rubriklerle puanlamadan sonra proje ve görevlerin analizinde yazılım araçlarının kullanıldığı görülmektedir. BİD becerisini değerlendirmeye özgü geliştirilmiş araçlar şunlardır: Dr. Scratch (Moreno-León ve Robles, 2015; Hoover vd., 2016; Romero vd., 2017; Anistyasari ve Kurniawan, 2018), Scratch analiz aracı (Chang vd., 2018), happyanalyzing.com (Brennan ve Resnick, 2012), semantic visual tool computational thinking pattern graph (Koh vd., 2010). Ayrıca Falloon (2015) videoları Video Analysis Software yardımıyla analiz etmektedir.

Yukarıdaki tabloda, veri analiz yöntemi belirtilmemiş veya açıklanmamış olan araştırmalar yer almamaktadır. Bu nedenle Tablo 3'teki herhangi bir veri toplama yöntemine ait frekans değeri ile Tablo 4'te bu yöntemle toplanan verilerin analizine ait frekans değeri farklılaşmaktadır. Bu açıdan incelendiğinde araştırmalarda özellikle görüşme ve gözlemden elde edilen verilerin analiz süreçlerine dair açıklamaların yetersiz kaldığı dikkat çekmektedir.

Veri Toplama Araçlarının İçerikleri

BİD becerisini değerlendirmek amacıyla kullanılan veri toplama araçlarının içeriklerine ait bulgular Tablo 5'te sunulmuştur.

Tablo 5. Veri Toplama Araçlarının İçerikleri

İçerik	f	Araştırma
Programlama	Geleneksel programlama	20 Brennan ve Resnick, 2012; Aiken vd., 2013; Seiter ve Foreman, 2013; Atmatzido ve Demetriadis, 2014; Falloon, 2015; Grover, 2015; Moreno-León ve Robles, 2015; Román-González, 2015; Basogain vd., 2016; Orton vd., 2016; Zhong vd., 2016; Grover vd., 2017; Romero vd., 2017; Zhang ve Biswas, 2017; Anistyasari ve Kurniawan, 2018; Chang vd., 2018; Hutchins vd., 2018; Looi vd., 2018; Pérez-Marín vd., 2018; Proctor ve Blikstein, 2018
	Oyun	11 Koh vd., 2010; Werner vd., 2012; Gouws vd., 2013a; Lee vd., 2014; Wang vd., 2014; Werner vd., 2014; Hoover vd., 2016; Rowe vd., 2017; Katchapakirin ve Anutariya, 2018; Leonard vd., 2016; Leonard vd., 2018
	Robotik	4 Bers vd., 2014; Atmatzidou ve Demetriadis, 2016; Chen vd., 2017; Leonard vd., 2018
Algı/tutum	4	Jun vd., 2017; Korkmaz vd., 2017; Weese ve Feldhausen, 2017; Leonard vd., 2018
Matematik	4	Yeh vd., 2011; Philip vd., 2013; Rowe vd., 2017; Bati vd., 2018
Günlük hayat	3	Dagiene ve Stupuriene, 2016; Chen vd., 2017; Witherspoon vd., 2017
Genel yetenek	2	Gouws vd., 2013b; Nesiba vd., 2015

Modelleme	2	Aiken vd., 2013; Hutchins vd., 2018
Standartlaştırılmış psikolojik testler	1	Kim vd., 2013
Bilgisayar bilimleri	1	Looi vd., 2018
Bilim	1	Bati vd., 2018
Diğer	2	Yeh vd., 2011 (tablolama yazılımı fonksiyonları); Philip vd., 2013 (veri yapıları)

Oldukça fazla sayıda araştırmanın BİD becerisini programlama içerikli ölçme araçlarıyla değerlendirmeyi amaçladığı, farklı disiplinlerin azınlıkta kaldığı Tablo 5'te görülmektedir. Ayrıca programlama içeriğine ait 35 ölçme aracından altısı bilim senaryolarını, biri matematik senaryosunu diğerleri ise; günlük hayat senaryolarını veya kodlama bilgilerini kullanmaktadır.

BİD becerisini değerlendirmede kullanılan standartlaştırılmış psikolojik test Roadranga, Yeany ve Padilla (1983) tarafından geliştirilmiş olan The Group Assessment of Logical Thinking (GALT)'dir.

Hedef Kitle

Değerlendirme yöntemlerinin hedef kitleye göre hazırlanmalıdır. Bu nedenle araştırmalar hitap ettiği seviye başlığında da analiz edilmiş ve bulgular Tablo 6'da paylaşılmıştır.

Tablo 6. Hedef Kitle

Hedef kitle	f	Araştırma
K-12	38	Koh vd., 2010; Brennan ve Resnick, 2012; Games ve Kane, 2012; Werner vd., 2012; Aiken vd., 2013; Seiter ve Foreman, 2013; Atmatzido ve Demetriadis, 2014; Bers vd., 2014; Lee vd., 2014; Wang vd., 2014; Werner vd., 2014; Falloon, 2015; Grover, 2015; Nesiba vd., 2015; Román-González, 2015; Atmatzidou ve Demetriadis, 2016; Basogain vd., 2016; Dagiene ve Stupuriene, 2016; Hoover vd., 2016; Leonard vd., 2016; Orton vd., 2016; Zhong vd., 2016; Chen vd., 2017; Grover vd., 2017; Jun vd., 2017; Rowe vd., 2017; Weese ve Feldhausen, 2017; Witherspoon vd., 2017; Zhang ve Biswas, 2017; Anistyasari ve Kurniawan, 2018; Bati vd., 2018; Hutchins vd., 2018; Katchapakirin ve Anutariya, 2018; Leonard vd., 2018; Ling vd., 2018; Looi vd., 2018; Pérez-Marín vd., 2018; Proctor ve Blikstein, 2018
Üniversite	7	Koh vd., 2010; Yeh vd., 2011; Gouws vd., 2013a; Gouws vd., 2013b; Philip vd., 2013; Korkmaz vd., 2017; Romero vd., 2017
Öğretmen	1	Kim vd., 2013
Her yaş	2	Moreno-León ve Robles, 2015; Chang vd., 2018

Tablo 6'da BİD becerisinin değerlendirildiği hedef kitle olarak okul öncesi, ilkökul, ortaokul ve lise seviyelerini kapsadığı için en fazla K-12 seviyesinin kullanıldığı görülmektedir. Üniversite düzeyindeki araştırmaların yapılan çalışmaların yaklaşık %15'ini oluşturmaktadır. Ayrıca bütün yaş gruplarını kapsayan değerlendirme yöntemleriyle ilgili iki araştırma bulunmaktadır.

Geçerlik ve Güvenirlik Çalışmaları

Araştırma sonuçlarının objektif ve kabul edilebilir olması için geçerli ve güvenilir ölçme araçlarının kullanılması gerekmektedir. Bu kapsamda incelenen BİD becerisini değerlendirme yöntemlerinden geçerlik veya güvenirlilik türlerinden en az bir tanesini içeren çalışmalara sahip olan araştırmalar Tablo 7'de yer almaktadır.

Tablo 7. Geçerlik ve Güvenirlik Çalışmaları

Geçerlik/Güvenirlik çalışması	f	Araştırma
Geçerlik	2	Koh vd., 2010; Dagiene ve Stupuriene, 2016
Güvenirlik	2	Philip vd., 2013; Weese ve Feldhausen, 2017
Hem geçerlik hem güvenilirlik	6	Román-González, 2015; Chen vd., 2017; Korkmaz vd., 2017; Witherspoon vd., 2017; Bati vd., 2018; Pérez-Marín vd., 2018

İncelenen değerlendirme yöntemlerinden %13 gibi oldukça düşük bir oranda araştırmanın hem geçerlik hem de güvenilirlik çalışmalarının bulunduğu tablo 7’de görülmektedir. Ayrıca yedi araştırmada proje, görev gibi verileri puanlamak için geliştirilen rubriklere ait güvenilirlik çalışmaları bulunmaktadır (Lee vd., 2014; Leonard vd., 2016; Chen vd., 2017; Grover vd., 2017; Leonard vd., 2018; Ling vd., 2018; Looi vd., 2018).

TARTIŞMA VE SONUÇ

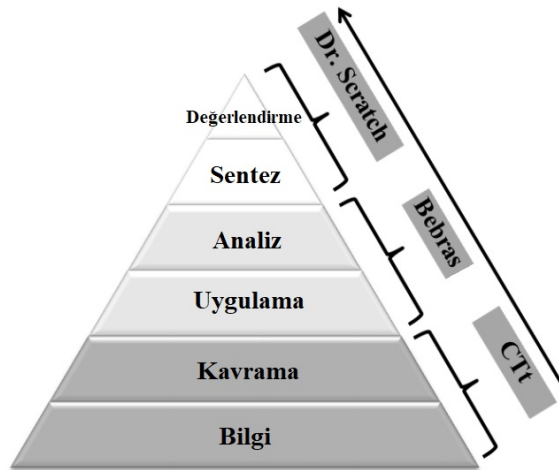
Dijital gelişmeler ışığında “teknolojiyi kullanarak problem çözme” olarak tanımlayabileceğimiz BİD becerisinin önemi de artmıştır (Gretter ve Yadav, 2016). Alanyazında bu becerinin kapsamıyla ilgili henüz fikir birliğine varılmamış olmakla beraber, geçerli ve güvenilir değerlendirme yöntemlerinin de oluşmadığı görülmektedir (Shute vd., 2017; Lockwood ve Mooney, 2018). Bu doğrultuda bu araştırmanın amacı alanyazında yer alan birbirinden farklı değerlendirme yöntemlerinin incelenmesi, eğilimin hangi yönde olduğunun belirlenmesi ve pozitif/negatif yönlerinin tartışılmasıdır.

İlk olarak BİD becerisinin kapsamının netleşmesine yardımcı olmak amacıyla en fazla değerlendirilen bileşenler araştırılmıştır. Elde edilen bulgular soyutlama, algoritmik düşünme, ayırıştırma, test etme hata ayıklama ve veri okuryazarlığının sıklıkla değerlendirilen ilk beş bileşen olduğunu ortaya koymuştur. De Araujo ve arkadaşları (2016) BİD becerisinin değerlendirilmesinde kullanılan yetenek/becerileri incelediklerinde, ilk üç bileşeni algoritmalar, soyutlama ve ayırıştırma olarak sıralamışlardır. Kalelioğlu ve arkadaşları (2016) tarafından yapılan alanyazın incelemesinde kavramsal çerçevede en fazla soyutlama, problem çözme, algoritmik düşünme ve örüntü tanıma bileşenlerinin yer aldığı belirlenmiştir. Hsu ve arkadaşlarının (2018) yaptıkları sistemik alanyazın çalışmasında BİD becerisini oluşturan bileşenler arasında en fazla soyutlama, algoritmik düşünme ve otomasyonun kullanıldığı ortaya koyulmuştur. Sondakh (2018) da BİD becerisini yükseköğretim düzeyinde değerlendirmek amacıyla araştırmacıların ağırlıklı olarak algoritmalar ve soyutlama bileşenlerini kullandıklarını ifade etmiştir. Bizim araştırmamızla benzer sonuçlara ulaşan çalışmaların yanısıra soyutlamanın BİD becerisinin en önemli bileşeni olduğu kabul gören kısımlardan biridir (Wing, 2006; Kramer, 2007; Grover ve Pea, 2013). Ayrıca araştırmacılar genellikle alanyazın incelemelerine dayanarak kendi seçtikleri bileşenleri değerlendirdiklerinin belirlenmesine karşın Brennan ve Resnick’in (2012) BİD becerisine ait ortaya koydukları kavramsal çerçevenin ve Dr. scratch analiz aracının bileşenlerinin farklı araştırmacılar tarafından da kabul görmeye başladığı tespit edilmiştir. Zhong ve arkadaşları (2016) da Brennan ve Resnick’in (2012) kavramsal çerçevesinin alanyazında dikkat çekici bir yere sahip olduğunu vurgulamışlardır.

İkinci olarak bu beceriye ait verilerin toplanma yöntemi incelenmiştir. Sırasıyla görev, çoktan seçmeli sorular, proje, açık uçlu sorular ve görüşme yöntemlerinin tercih edildiği belirlenmiştir. Yine De Araujo ve arkadaşları (2016) BİD becerisini değerlendirmek amacıyla en

fazla çoktan seçmeli soruların kullanıldığını ifade etmektedirler. Weinberg (2013) doktora tezinde bu becerinin değerlendirilmesinde en fazla anketlerden, öğretmen veya araştırmacı tarafından geliştirilen testlerden ve öğrenci çalışmalarından veri toplandığını ortaya koymuştur. Brennan ve Resnick (2012) görüşme ana başlığı altında bulunan görüşme türlerinin avantajlı ve dezavantajlı yönleriyle ilgili açıklamalar yapmışlardır. Ürün tabanlı görüşmenin süreç odaklı incelemeler yapma fırsatı sunarken, zaman gerektirmesi ve öğrencilerin önceden oluşturdukları ürünler hakkında görüşüldüğü için hatırlama sorunu yaşanabilmesi gibi dezavantajları bulunduğunu ifade etmişlerdir. Bu sınırlılıkların giderilmesi için hafızaya dayalı olmayana tasarım senaryolarıyla eylem esnasında gözlem yapılabileceği belirtilmiştir. Fakat bu yöntemde de senaryoların öğrencilerin ilgisini çekmemesi ve motivasyonlarını kaybetmeleri veya süreç içerisinde tıkanıp ilerleme gösteremeyebileceklerinden bahsedilmiştir. Ayrıca Grover (2015) BİD becerisi gibi karmaşık yapılar için “değerlendirmeler sistemi” olarak adlandırdığı farklı disiplinlerde, birden fazla veri toplama yönteminin kullanılmasını önermektedir. Fakat araştırmanın sonuçları araştırmacıların hala tek veri toplama yöntemini tercih ettiklerini göstermektedir.

Üçüncü olarak toplanan verilerin analiz yöntemleri araştırılmıştır. Bulgular likert veya rubrikle puanlama yönteminin toplanan verilerin analiz edilmesinde en fazla kullanılan yöntem olduğunu ortaya koymuştur. Ling ve arkadaşları (2018) tarafından geliştirilen rubrik sayesinde farklı yöntemlerle toplanmış veriler BİD becerisi kapsamında analiz edilebilmektedir. Araştırmacıların belirledikleri kriterlere göre yaptıkları puanlamaların yanısıra bazı yazılımsal araçlarının geliştirildiği belirlenmiştir. Román-González, Moreno-León ve Robles (2017) BİD becerisi kapsamında proje ve görevlerin analizinde kullanılan Dr. Scratch aracının veri toplama araçları olan Bebras görevleri ve CTt ile olan ilişkisini Bloom taksonomisinin seviyeleri üzerinde açıklamışlardır. Bu ilişki Şekil 3’te görülmektedir.



Şekil 3. BİD becerisinin değerlendirilmesinde kullanılan ölçme araçları ve Bloom taksonomisi (Román-González vd., 2017)

Çalışmanın; BİD becerisini ölçmek için yazılım araçlarının kullanılmasını savunan görüşlere ilişkin sonuçlarına karşın, proje ve görevlerin analizinde kullanılan Dr. Scratch gibi yazılım araçlarının teknik olarak karmaşıklığı ölçtüğünü, süreci ve oluşturulan ürünün yaratıcılığını veya amacına hizmet edip etmediğini analiz etmekte yetersiz kaldığını ifade eden görüşler bulunmaktadır. (Romero vd., 2017). Hoover ve arkadaşları (2016) kullanışsız ürünlerin uygun bloklarla oluşturulması durumunda blok temelli araçların bu durumu

yakalayamayacağını, bu araçların programlamaya ait süreçlerin değerlendirilmesi amacıyla kullanılmasının daha uygun olacağını belirtmişlerdir. Araştırmanın sonuçları Bebras görevlerinin BİD becerisini değerlendirmek için uygun bir araç olduğunu belirtmesine karşın, Araujo, Santos, Andrade, Guerrero ve Dagiené (2017) BİD becerisini değerlendirmek amacıyla lisans düzeyinde derledikleri Bebras görevlerinden oluşan bir testi kullandıkları çalışmalarında, bu yöntemin BİD becerisini değerlendirmek için uygun görünmediğini ifade etmişlerdir.

Çalışmanın diğer bir kaydedeğer sonucu, BİD becerisinin değerlendirilmesi amacıyla kullanılan araçların çok büyük bir oranda programlama içeriğini kullandığını göstermektedir. Yapılan sistematik alanyazın inceleme çalışmalarında BİD becerisini geliştirmeye yönelik en fazla programlama içeriğinin tercih edildiği ortaya koyulmuştur (De Araujo vd., 2016; Lockwood ve Mooney, 2018). Yine Hsu ve arkadaşları (2018) alanyazın incelemelerinde BİD becerisini geliştirmek amacıyla büyük oranda programlama ve bilgisayar bilimleri içeriklerinin kullanıldığını göstermiştir. Buna karşılık Wing (2006) BİD becerisinin yalnızca bilgisayar bilimcilerin veya mühendislerin değil, yeni nesil bireylerde bulunması gereken bir beceri olduğunu vurgulamıştır. Araştırmacıların bu beceriyi ağırlıklı olarak programlama ve bilgisayar bilimleri içerikleriyle geliştirme ve değerlendirme çalışmaları alanyazının çelişkili yanı olarak görülebilir.

BİD becerisini değerlendirme yöntemlerinin hedef kitlesi incelendiğinde sırasıyla K-12 öğrencilerinin, üniversite öğrencilerinin ve öğretmenlerin yer aldığı görülmektedir. Üniversite öğrencileri, öğretmenler veya yetişkinler gibi farklı düzeyde değerlendirmelerin oldukça az sayıda olduğu görülmektedir. De Araujo ve arkadaşları (2016) da benzer şekilde K-12, üniversite, öğretmen ve okul öncesi seviyelerinde araştırmaların yürütüldüğünü ortaya koymuştur. Kalelioğlu ve arkadaşları (2016) BİD becerisi ile ilgili araştırmaları inceledikleri çalışmalarında K-12 ve yükseköğretim düzeyinde yürütülen araştırmaların sayılarının eşit düzeyde olduğunu ifade etmişlerdir. Bu farklı sonucun yükseköğretim düzeyinde BİD becerisini konu alan araştırmalarda, bu becerinin değerlendirilmesini de kapsayan çalışmaların %40 oranında kalmasından dolayı ortaya çıktığı düşünülmektedir (Sondakh, 2018).

Son olarak BİD becerisini değerlendirme yöntemleri ile ilgili araştırmaların %4'ünde geçerlik, %4'ünde güvenilirlik ve %13'ünde hem geçerlik hem de güvenilirliğe dair çalışmaların yer aldığı belirlenmiştir. Bu bulgu, alanyazında BİD becerisinin değerlendirilmesi için kabul edilen yöntemlerin oluşmadığına ve geçerli ve güvenilir ölçme araçlarına ihtiyaç duyulduğuna dair hem fikir olunan görüşle uyusmaktadır (Grover, 2015; Shute vd., 2017; Lockwood ve Mooney, 2018).

ÖNERİLER

Araştırmadan elde edilen bulgular ışığında BİD becerisini ölçmeyi amaçlayan öğretmen ve araştırmacılara aşağıdaki önerilerde bulunmaktadır.

1. Çalışmada elde edilen sonuçlardan birisi, Grover'in (2015) da ifade ettiği, karmaşık üst düzey becerilerin değerlendirilmesi için birden fazla veri toplama yönteminin kullanılmasına dair önerisine karşın, araştırmaların büyük çoğunluğunda BİD becerisini değerlendirmek amacıyla ağırlıklı olarak tek veri toplama yönteminin kullanıldığını göstermektedir. Dolayısıyla gelecek araştırmalarda, BİD becerisine dair farklı veri toplama yöntemlerinin bir arada kullanıldığı çalışmaların planlanması önerilmektedir.

2. Çalışmanın en kaydedeğer bulgularından birisi olarak elde edilen; araştırmalarda toplanan verilerin analiz yöntemleri incelenmesi sonucunda, özellikle gözlem ve görüşme yoluyla

toplanan verilerin analizleri hakkında açıklamaların bulunmaması dikkat çekmektedir. Bu verilerin analiz yöntemleri hakkında açıklamaların daha detaylı raporlanması ve nitel tartışmalara yer verilmesi alana katkı sağlayacaktır.

3. Çalışmanın sonuçları, BİD becerisini değerlendirme yöntemlerinin içeriklerinin yaklaşık %63'ünün programlamadan oluştuğunu ortaya koymaktadır. Oysaki bu becerinin bilgisayar bilimciler veya mühendislerin değil yeni nesil bireylerin sahip olması gereken becerilerden biri olduğuna dair birçok görüş bulunmaktadır (Wing, 2006; Barr, Harrison ve Conery, 2011; Gretter ve Yadav, 2016). Ayrıca Grover (2015)'in önerdiği "değerlendirmeler sisteminde" farklı disiplinlerde içeriklerin kullanılmasından bahsedilmektedir. Bu nedenlerle ölçme aracı geliştirecek öğretmen veya araştırmacıların, programlamanın yanısıra farklı disiplinlerde içerikleri kullanmaları faydalı olacaktır.

4. Araştırmanın bulguları BİD becerisinin %79 gibi büyük bir oranda K-12 düzeyinde değerlendirildiğini göstermektedir. Bu becerinin değerlendirilmesine yönelik olarak yapılacak araştırmalarda yetişkin grupların tercih edilmesinin alana katkı sağlayacağı düşünülmektedir.

5. Son olarak, elde edilen bulgular, BİD becerisinin değerlendirilmesiyle ilgili yöntemlerin geçerlik ve güvenilirlik çalışmalarının yetersiz olmasından dolayı öneri seviyesinde kaldığını göstermektedir. Gelecek araştırmaların alanyazında yer alan mevcut yöntemlerin geçerlik ve güvenilirlik çalışmalarına devam etmeleri veya kendi geliştirdikleri yöntemlerin geçerlik ve güvenilirlik bölümleriyle ilgili bölümlerine hassasiyet göstermeleri tavsiye edilmektedir.

KAYNAKÇA

- Aiken, J. M., Caballero, M. D., Douglas, S. S., Burk, J. B., Scanlon, E. M., Thoms, B. D., et. al. (2013). Understanding student computational thinking with computational modeling. *Proceedings of American Institute of Physics Conference, 1513* (1), 46-49.
- Andrews, R., & Harlen, W. (2006). Issues in synthesizing research in education. *Educational Research, 48*(3), 287-299.
- Anistyasari, Y., & Kurniawan, A. (2018). Exploring Computational Thinking to Improve Energy-Efficient Programming Skills. *Proceedings of MATEC Web of Conferences, 197*, 15011.
- Araujo, A. L. S. O., Santos, J. S., Andrade, W. L., Guerrero, D. D. S., & Dagièné, V. (2017). Exploring computational thinking assessment in introductory programming courses. *Proceedings of Frontiers in Education Conference, 1-9*.
- Atmatzidou, S., & Demetriadis, S. (2014). How to support students' computational thinking skills in educational robotics activities. *Proceedings of Forth International Workshop Teaching Robotics, Teaching with Robotics & Fifth International Conference Robotics in Education, 43-50*.
- Atmatzidou, S., & Demetriadis, S. (2016). Advancing students' computational thinking skills through educational robotics: A study on age and gender relevant differences. *Robotics and Autonomous Systems, 75*, 661-670.
- Bakırcı, H., & Kutlu, E. (2018). Fen bilimleri öğretmenlerinin FeTeMM yaklaşımı hakkındaki görüşlerinin belirlenmesi. *Turkish Journal of Computer and Mathematics Education, 9*(2), 367-389.
- Barr, D., Harrison, J., & Conery, L. (2011). Computational thinking: A digital age skill for everyone. *Learning & Leading with Technology, 38*(6), 20-23.
- Basogain, X., Olabe, M. A., Olabe, J. C., Ramírez, R., Del Rosario, M., & Garcia, J. (2016). PC-01: Introduction to computational thinking: Educational technology in primary and secondary education. *Proceeding of*

International Symposium on Computers in Education, 1-5.

- Bati, K., Yetişir, M. I., Çalışkan, I., Güneş, G., & Gül Saçan, E. (2018). Teaching the concept of time: A steam-based program on computational thinking in science education. *Cogent Education*, 5(1), 1-16.
- Bell, T., Alexander, J., Freeman, I., & Grimley, M. (2009). Computer science unplugged: School students doing real computing without computers. *The New Zealand Journal of Applied Computing and Information Technology*, 13(1), 20-29.
- Bers, M. U., Flannery, L., Kazakoff, E. R., & Sullivan, A. (2014). Computational thinking and tinkering: Exploration of an early childhood robotics curriculum. *Computers & Education*, 72, 145-157.
- Brennan, K., & Resnick, M. (2012). *New frameworks for studying and assessing the development of computational thinking*. Paper presented at the Annual Meeting of the American Educational Research Association, Vancouver, Canada.
- Caldwell, H., & Smith, N. (Eds.). (2016). *Teaching Computing Unplugged in Primary Schools: Exploring Primary Computing Through Practical Activities Away from the Computer*. London: Sage Publication.
- Chang, Z., Sun, Y., Wu, T. Y., & Guizani, M. (2018). Scratch analysis tool (SAT): A modern scratch project analysis tool based on ANTLR to assess computational thinking skills. *Proceedings of the 14th International Wireless Communications & Mobile Computing Conference*, 950-955.
- Chen, G., Shen, J., Barth-Cohen, L., Jiang, S., Huang, X., & Eltoukhy, M. (2017). Assessing elementary students' computational thinking in everyday reasoning and robotics programming. *Computers & Education*, 109, 162-175.
- Dagiene, V., & Stupuriene, G. (2016). Bebras--A sustainable community building model for the concept based learning of informatics and computational thinking. *Informatics in Education*, 15(1), 25-44.
- De Araujo, A. L. S. O., Andrade, W. L., & Guerrero, D. D. S. (2016). A systematic mapping study on assessing computational thinking abilities. *Proceedings of Frontiers in Education Conference*, 1-9.
- Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107-115.
- Ergai, A., Cohen, T., Sharp, J., Wiegmann, D., Gramopadhye, A., & Shappell, S. (2016). Assessment of the human factors analysis and classification system (HFACS): Intra-rater and inter-rater reliability. *Safety Science*, 82, 393-398.
- Falloon, G. (2015). Building computational thinking through programming in K-6 education: A New Zealand experience. In L. Gomez Chova, A. Lopez Martinez, & I. Chandel Torres (Eds.), *EDULearn zProceedings* (pp. 882-892). Barcelona, Spain: IATED Academy.
- Games, A., & Kane, L. (2012). *Examining trends in adolescents' computational thinking skills within the globaloria educational game design environment*. Retrieved January 8, 2019, from <http://www.worldwideworkshop.org/pdfs/GlobaloriaExaminingTrendsAdolescentsComputSkillsGamesKaneAug2012.pdf>
- Gouws, L. A., Bradshaw, K., & Wentworth, P. (2013a). Computational thinking in educational activities: an evaluation of the educational game light-bot. *Proceedings of the 18th ACM Conference on Innovation and Technology in Computer Science Education*, 10-15.
- Gouws, L. A., Bradshaw, K., & Wentworth, P. (2013b). First year student performance in a test for computational thinking. *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference*, 271-277.
- Gretter, S., & Yadav, A. (2016). Computational thinking and media & information literacy: An integrated approach to teaching twenty-first century skills. *TechTrends*, 60(5), 510-516.

- Grover, S. (2015). "Systems of Assessments" for deeper learning of computational thinking in K-12. *Proceedings of the 2015 Annual Meeting of the American Educational Research Association*, 15-20.
- Grover, S., Bienkowski, M., Basu, S., Eagle, M., Diana, N., & Stamper, J. (2017). A framework for hypothesis-driven approaches to support data-driven learning analytics in measuring computational thinking in block-based programming. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 530-531.
- Grover, S., & Pea, R. (2013). Computational thinking in K-12: A review of the state of the field. *Educational Researcher*, 42(1), 38-43.
- Guba, E. G., & Lincoln, Y. S. (1982). Epistemological and methodological bases of naturalistic inquiry. *Educational Communication and Technology Journal*, 30 (4), 233-252
- Hoover, A. K., Barnes, J., Fatehi, B., Moreno-León, J., Puttick, G., Tucker-Raymond, E., et. al. (2016). Assessing computational thinking in students' game designs. *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, 173-179.
- Hsu, T. C., Chang, S. C., & Hung, Y. T. (2018). How to learn and how to teach computational thinking: Suggestions based on a review of the literature. *Computers & Education*, 126, 296-310.
- Hutchins, N., Biswas, G., Conlin, L., Emara, M., Grover, S., Basu, S., et. al. (2018). *Studying synergistic learning of physics and computational thinking in a learning by modeling environment*. Paper presented at the 26th International Conference on Computers in Education, Manila, Philippines.
- ISTE. (2014). *Computational thinking for all*. Retrieved April 18, 2019, from <https://www.iste.org/explore/articleDetail?articleid=152&category=Solutions&article=Computational-thinking-for-all>
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130-144.
- Jun, S., Han, S., & Kim, S. (2017). Effect of design-based learning on improving computational thinking. *Behaviour & Information Technology*, 36(1), 43-53.
- Karaçam, Z. (2013). Sistematik derleme metodolojisi: Sistematik derleme hazırlamak için bir rehber. *Dokuz Eylül Üniversitesi Hemşirelik Fakültesi Elektronik Dergisi*, 6(1), 26-33.
- Kalelioglu, F., Gülbahar, Y., & Kukul, V. (2016). A framework for computational thinking based on a systematic research review. *Baltic Journal of Modern Computing*, 4(3), 583-596.
- Katchapakirin, K., & Anutariya, C. (2018). *An architectural design of ScratchThAI: A conversational agent for computational thinking development using Scratch*. Paper presented at the 10th International Conference on Advances in Information Technology, Bangkok, Thailand.
- Kim, B., Kim, T., & Kim, J. (2013). Paper and pencil programming strategy toward computational thinking for non-majors: Design your solution. *Journal of Educational Computing Research*, 49(4), 437-459.
- Kitchenham, B. (2004). *Procedures for performing systematic reviews*. (NICTA Technical Report 0400011T.1). Keele, UK: Keele University Department of Computer Science.
- Koh, K. H., Basawapatna, A., Bennett, V., & Repenning, A. (2010). Towards the automatic recognition of computational thinking for adaptive visual language learning. *Proceedings of Symposium on Visual Languages and Human-Centric Computing*, 59-66.
- Korkmaz, Ö., Çakır, R., & Özden, M. Y. (2017). A validity and reliability study of the computational thinking scales (CTS). *Computers in Human Behavior*, 72, 558-569.
- Kramer, J. (2007). Is abstraction the key to computing?. *Communications of the ACM*, 50(4), 36-42.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*,

33(1), 159-174.

- Lee, T. Y., Mauriello, M. L., Ahn, J., & Bederson, B. B. (2014). CTArcade: Computational thinking with games in school age children. *International Journal of Child-Computer Interaction*, 2(1), 26-33.
- Leonard, J., Buss, A., Gamboa, R., Mitchell, M., Fashola, O. S., Hubert, T., et. al. (2016). Using robotics and game design to enhance children's self-efficacy, STEM attitudes, and computational thinking skills. *Journal of Science Education and Technology*, 25(6), 860-876.
- Leonard, J., Mitchell, M., Barnes-Johnson, J., Unertl, A., Outka-Hill, J., Robinson, R., et. al. (2018). Preparing teachers to engage rural students in computational thinking through robotics, game design, and culturally responsive teaching. *Journal of Teacher Education*, 69(4), 386-407.
- Ling, U. L., Saibin, T. C., Naharu, N., Labadin, J., & Aziz, N. A. (2018). An evaluation tool to measure computational thinking skills: pilot investigation. *National Academy of Managerial Staff of Culture and Arts Herald*, 1, 606-614.
- Lockwood, J., & Mooney, A. (2018). Computational thinking in secondary education: Where does it fit? A systematic literary review. *International Journal of Computer Science Education in Schools*, 2(1), 41-60.
- Looi, C. K., How, M. L., Longkai, W., Seow, P., & Liu, L. (2018). Analysis of linkages between an unplugged activity and the development of computational thinking. *Computer Science Education*, 28(3), 255-279.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 22(3), 276-282.
- Moreno-León, J., & Robles, G. (2015). *Analyze your Scratch projects with Dr. Scratch and assess your computational thinking skills*. Paper presented at the Scratch Conference, Amsterdam.
- Román-González, M., Moreno-León, J., & Robles, G. (2017). Complementary tools for computational thinking assessment. *Proceedings of International Conference on Computational Thinking Education*, 154-159.
- Nesiba, N., Pontelli, E., & Staley, T. (2015, October). DISSECT: Exploring the relationship between computational thinking and English literature in K-12 curricula. *Proceedings of Frontiers in Education Conference*, 1-8.
- OECD. (2018). *The future of education and skills: Education 2030*. Retrieved July 22, 2019, from [https://www.oecd.org/education/2030/E2030%20Position%20Paper%20\(05.04.2018\).pdf](https://www.oecd.org/education/2030/E2030%20Position%20Paper%20(05.04.2018).pdf)
- Orton, K., Weintrop, D., Beheshti, E., Horn, M., Jona, K., & Wilensky, U. (2016). Bringing computational thinking into high school mathematics and science classrooms. *Proceedings of 12th International Conference of the Learning Sciences*, 2, 705-712.
- Pérez-Marín, D., Hijón-Neira, R., Bacelo, A., & Pizarro, C. (2018). Can computational thinking be improved by using a methodology based on metaphors and scratch to teach computer programming to children?. *Computers in Human Behavior*, (in press).
- Philip, M., Renumol, V. G., & Gopeekrishnan, R. (2013). A pragmatic approach to develop computational thinking skills in novices in computing education. *Proceedings of International Conference in MOOC, Innovation and Technology in Education*, 199-204.
- Proctor, C., & Blikstein, P. (2018). How Broad is Computational Thinking? A Longitudinal Study of Practices Shaping Learning in Computer Science. *Proceedings of the 13th International Conference of the Learning Sciences*, 1, 544-551.
- Roadrangka, V., Yeany, R. H., & Padilla, M. J. (1983). *The construction and validation of group assessment of logical thinking (GALT)*. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, Dallas, USA.
- Román-González, M. (2015). Computational thinking test: Design guidelines and content validation.

Proceedings of EDULEARN15 Conference, 2436-2444.

- Romero, M., Lepage, A., & Lille, B. (2017). Computational thinking development through creative programming in higher education. *International Journal of Educational Technology in Higher Education, 14*(42), 1-15.
- Rowe, E., Asbell-Clarke, J., Cunningham, K., & Gasca, S. (2017). Assessing implicit computational thinking in Zoombinis Gameplay: Pizza Pass, Fleens & Bubblewonder Abyss. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play, 195-200.*
- Seiter, L., & Foreman, B. (2013). Modeling the learning progressions of computational thinking of primary grade students. *Proceedings of the Ninth Annual International ACM Conference on International Computing Education Research, 59-66.*
- Sengupta, P., Kinnebrew, J. S., Basu, S., Biswas, G., & Clark, D. (2013). Integrating computational thinking with K-12 science education using agent-based computation: A theoretical framework. *Education and Information Technologies, 18*(2), 351-380.
- Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review, 22*, 142-158.
- Sondakh, D. E. (2018). *Review of computational thinking assessment in higher education*. Retrieved May 29, 2019, from https://www.researchgate.net/profile/Debby_Sondakh/publication/324984840_Review_of_Computational_Thinking_Assessment_in_Higher_Education/links/5af0378aa6fdcc8508b96527/Review-of-Computational-Thinking-Assessment-in-Higher-Education.pdf
- Streubert, H. J., & Carpenter, D. R. (2011). *Qualitative research in nursing*. (5th Ed.). Philadelphia: Lippincott Williams ve Wilkins.
- Wang, D., Wang, T., & Liu, Z. (2014). A tangible programming tool for children to cultivate computational thinking. *The Scientific World Journal, 2014*, 1-11.
- Weese, J. L., & Feldhausen, R. (2017). *STEM Outreach: Assessing computational thinking and problem solving*. Paper presented at Annual American Society for Engineering Education Conference & Exposition, Ohio, USA.
- Weinberg, A. E. (2013). *Computational thinking: An investigation of the existing scholarship and research*. Unpublished Doctoral Dissertation, Colorado State University, Colorado, USA.
- Werner, L., Denner, J., & Campe, S. (2014). Using computer game programming to teach computational thinking skills. In K. Schrier (Ed.), *Learning, Education and Games* (pp.37-53). Pittsburgh: ETC Press.
- Werner, L., Denner, J., Campe, S., & Kawamoto, D. C. (2012). The fairy performance assessment: measuring computational thinking in middle school. *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education, 215-220.*
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33-35.
- Witherspoon, E. B., Higashi, R. M., Schunn, C. D., Baehr, E. C., & Shoop, R. (2017). Developing computational thinking through a virtual robotics programming curriculum. *ACM Transactions on Computing Education, 18*(1), 1-20.
- Yağcı, M. (2019). A valid and reliable tool for examining computational thinking skills. *Education and Information Technologies, 24*(1), 929-951.
- Yeh, K. C., Xie, Y., & Ke, F. (2011). Teaching computational thinking to non-computing majors using spreadsheet functions. *Proceedings of Frontiers in Education Conference, 1-5.*
- Zhang, N., & Biswas, G. (2017). Assessing students computational thinking in a learning by modeling

environment. *Proceedings of International Conference on Computational Thinking Education*, 11-16.

Zhong, B., Wang, Q., Chen, J., & Li, Y. (2016). An exploration of three-dimensional integrated assessment for computational thinking. *Journal of Educational Computing Research*, 53(4), 562-590.

EXTENDED SUMMARY

INTRODUCTION

Computational thinking (CT) has been added to the list of expected skills from individuals in parallel with technological advances. Researchers have developed many content in order to improve CT which can be defined as “problem solving with using technology”. It is noteworthy that in the literature, there are no fully-accepted, valid, and reliable methods to evaluate the development of this skill. Therefore, identifying the different assessment methods, revealing their differences, and discussing the positive/negative aspects of these methods will contribute to the new assessment methods to be developed.

METHOD

In this research, systematic literature review was used. Systematic literature review is the process of thorough scanning of the research related to the research question, then exclusion according to the criteria and obtaining synthesis from the selected studies (Kitchenham, 2004).

Scanning Process and Selected Research

While determining the studies to be examined in this process, Web of Science (WOS), ERIC and Scholar Google databases were searched. Scanning was done in the titles of the studies which published by 2018 and only “computational thinking” was chosen as the keyword. 1621 studies (518 in WOS, 106 in ERIC, 997 in Scholar Google) were examined according to inclusion criteria. 157 studies that meet the criterias were selected for detailed reading. After the detailed reading process, a total of 110 studies were excluded according to the exclusion criterias. As a result, 47 studies were selected to analyze.

Data Analysis, Validity and Reliability

A researcher conducted content analysis for 47 studies. The researcher examined each study in detail under the headings of data collection method, content of the measurement tool, conceptual framework, target audience and validity/reliability. In this process, she took notes from the headings mentioned, and then compared these notes with each other, themed similar ones and created new themes for differentiating situations (Elo and Kyngäs, 2008). Consistency was examined to increase internal reliability (Guba and Lincoln, 1982). For this purpose, 47 studies were re-read by the same researcher two months later to ensure that themes were consistent with all studies and to ensure intra-rater reliability (Ergai et al., 2016). Cohen Kappa coefficient was tested for compatibility between the two codes (McHugh, 2012). According to Landis and Koch (1977), the obtained 0.82 value is between 0.81 and 1.00, which shows that there is a very good level of compliance. In order to increase the external reliability by providing confirmability, the tables of the findings contain the information of the research (Streubert and Carpenter, 2011). In addition, all processes of the research to increase external validity are described in detail (Guba and Lincoln, 1982).

FINDINGS

As a result of the analysis, the findings related to conceptual framework, data collection method, data analysis method, content of measurement instrument, target audience and validity/reliability are as follows:

1- Conceptual framework: The most commonly used components are listed as abstraction, algorithmic thinking, decomposition, testing and debugging, data literacy, sequencing, flow control structures, logical thinking, parallelization, etc.

2- Data collection method: According to the frequency of preference, the data collection methods used were task, multiple choice question, project, open-ended question and interview. System records, observation and likert-type questionnaires were less frequently used. Multiple data collection methods were used in 22 studies and single data collection method was used in 24 studies in order to evaluate CT.

3- Data analysis method: The results found out that likert or rubrics scoring method were more preferred than software tools in the analysis of projects and tasks. Specific tools for assessing CT include: Scratch (Moreno-León and Robles, 2015; Anistiyasari and Kurniawan, 2018), Scratch analysis tool (Chang et

al., 2018), happyanalyzing.com (Brennan and Resnick, 2012), semantic visual tool computational thinking pattern graph (Koh et al., 2010).

4- The content of the measurement tool: The findings showed that 63% of the programming content was used and that different disciplines such as perception/attitude, mathematics and general talent remained in the minority.

5- Target group: The K-12 level, which covers preschool, primary, secondary and high school levels, is frequently preferred. University level research constitutes approximately 15% of the studies.

6- Validity and reliability: Two research have validity studies and 3 research have reliability studies. Only 6 research have both validity and reliability studies

DISCUSSION AND CONCLUSION

Firstly, the most evaluated components were investigated in order to contribute to clarifying the scope of CT and the results were similar to the findings of the literature. (De Araujo et al., 2016; Kalelioglu et al., 2016; Sondakh, 2018)

The findings that the most preferred data collection methods are tasks, multiple choice questions, project, open-ended questions and interview are in agreement with the literature (Weinberg, 2013; De Araujo et al., 2016). The types of interviews used eliminate the limitations of each other (Brennan and Resnick, 2012). Despite Grover's (2015) recommendation to use more than one method to assess CT, results found out that data was still collected in one method.

When the analysis of the collected data was examined, findings showed that software tools specific to CT were used. However, there are criticisms that these tools will not be sufficient to measure CT (Hoover vd., 2016; Romero vd., 2017).

The results that the measurement tools have a very large proportion of programming content are similar to those in the literature (De Araujo et al., 2016; Lockwood and Mooney, 2018). Although Wing (2006) emphasizes that CT is a skill that should be found not only by computer scientists or engineers but also by new generation individuals, it can be seen as a contradictory aspect of the literature that the researchers mainly evaluate this skill with programming and computer science contents.

De Araujo et al. (2016) reported that more studies were conducted at the K-12 level, similar to the results of this study.

Finally, there are 6 out of 47 studies which include both validity and reliability were obtained. that in the literature, This finding is in agreement with the opinion that there are no fully-accepted, valid, and reliable methods to evaluate the development of this skill (Shute vd., 2017; Lockwood ve Mooney, 2018).