

## PAPER DETAILS

TITLE: Zumbo`nun madde tepki sürecine yönelik elestirel bakis açisiyla Türkçeyi Ölçme ve degerlendirmede yeni yaklasimlar

AUTHORS: Talha GÖKTENTÜRK,Mehmet Hilmi SAGLAM,Bruno D ZUMBO

PAGES: 224-245

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/3092845>

**15. Zumbo'nun madde tepki sürecine yönelik eleştirel bakış açısıyla Türkçeyi ölçme ve değerlendirmede yeni yaklaşımlar****Talha GÖKTENTÜRK<sup>1</sup>****Mehmet Hilmi SAĞLAM<sup>2</sup>****Bruno D. ZUMBO<sup>3</sup>**

**APA:** Göktentürk, T. & Sağlam, M. H. & Zumbo, B. D. (2023). Zumbo'nun madde tepki sürecine yönelik eleştirel bakış açısıyla Türkçeyi ölçme ve değerlendirmede yeni yaklaşımlar. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi*, (33), 223-245. DOI: 10.29000/rumelide.1285296

**Öz**

Ana dili olarak Türkçenin ölçülmesi ve değerlendirilmesi maddenin kendisi ve maddeye yönelik katılımcı tepkilerine odaklı bir yaklaşımla sürmektedir. 21. yüzyılda Türkçenin ölçülmesi ise MEB ve ÖSYM öncülüğündeki testler yoluyla gerçekleşmekte ve aynı yaklaşımla paydaşlar tarafından yürütülmektedir. Bununla beraber Dünya'da maddeye tepki sürecinin ekolojik arka planının araştırılması önemini giderek artırmaktadır. Madde tepki sürecinde madde ile karşılaşan katılımcı, tepkisinde arka plandaki kompleks bir ekolojik örüntüden etkilenmektedir. Dolayısıyla ilgili örüntünün geçerlilik araştırmalarına dâhil edilmemesi tablonun sınırlı görülebilmesine ve ölçme sonuçlarının eksik yorumlanmasına neden olmaktadır. Madde yanlılığı araştırmalarının mühim bir parçası olan diferansiyel madde fonksiyonunu belirlemeye yönelik çalışmalar bu noktada umut vermektedir. Buradan hareketle Türkçenin ana dili olarak ölçülmesindeki kısıtlı yaklaşımlara da karşılık 3 adımda bu çalışma şekillendirilmiştir. Birinci adımda ana dili olarak Türkçenin ölçülmesi ve değerlendirilmesine yönelik bir çerçeve sunulmuştur. İkinci adımda madde tepki sürecinin ekolojisi tartışılmıştır. Üçüncü adımda kompleks arka planın keşfi için diferansiyel madde fonksiyonu araştırmalarından potansiyel bir çözüm olarak Zumbo ve diğerleri (2015) tarafından geliştirilen ekolojik model tanıtılmıştır. Son aşamada ise kurum bazında MEB ve ÖSYM başta olmak üzere madde yazarları, araştırmacılar ve politika yapımcılar için tavsiye ve teklifler sıralanmıştır. Tartışmanın çıkarımından hareketle ana dili olarak Türkçenin ölçülmesi ve değerlendirilmesinde ekolojik yaklaşımın ve ekolojik modelin geçerlilik araştırmaları için büyük bir potansiyel taşıdığı söylenebilir. Dolayısıyla Türkiye'nin ölçme ihtiyaçlarına uygun olarak ekoloji tabanlı bir perspektifin Türkçenin ana dilinde ölçülmesine adapte edilmesi gerekmektedir. Böylece Türkçe dersi öğretim programının ölçmeye ve değerlendirmeye yönelik yaklaşım ihtiyacının karşılanması mümkün olacaktır.

**Anahtar kelimeler:** maddeye tepki süreci, ölçme ve değerlendirme, diferansiyel madde fonksiyonu, Türkçe, okuma, ekolojik model

<sup>1</sup> Dr. Öğr. Üyesi, Yıldız Teknik Üniversitesi, Eğitim Fakültesi, Türkçe ve Sosyal Bilimler Eğitimi Bölümü, Türkçe Eğitimi ABD, (İstanbul, Türkiye), talhagoktenturk@gmail.com, ORCID ID: 0000-0002-8619-0698 [Araştırma makalesi, Makale kayıt tarihi: 18.03.2023 kabul tarihi: 20.04.2023; DOI: 10.29000/rumelide.1285296]

<sup>2</sup> Arş. Gör., The University of British Columbia, Faculty of Education, The Department of Educational and Counselling Psychology, and Special Education (Vancouver, BC, Kanada), mehmethilmissaglam@gmail.com, ORCID ID: 0000-0001-5746-738X

<sup>3</sup> Prof. Dr., The University of British Columbia, Faculty of Education, The Department of Educational and Counselling Psychology, and Special Education (Vancouver, BC, Kanada), bruno.zumbo@ubc.ca, ORCID ID: 0000-0003-2885-5724

## New lens in assessment and evaluation of Turkish language with Zumbo's critical view of item responding

### Abstract

The measurement and evaluation of the Turkish language continue with an approach focused on the item and participant responses. In the 21st century, the measurement of Turkish is being made through tests led by ÖSYM and is carried out by stakeholders with the same approach. However, a new framework on the ecological background of the item-responding process is increasingly important worldwide. From this point of view, the respondent is affected by a complex ecological background reflected in the item response process. Therefore, leaving the relevant pattern outside of assessment design and validity studies gives only a limited view of test results and can cause the misinterpretation of measurement findings. Literature on differential item function, an important part of item bias research, gives hope at this point. In response to the approaches to measuring Turkish as a first language (i.e., native language or mother tongue), this study was shaped in 3 steps. The first step presented a framework for measuring and evaluating Turkish. In the second step, the ecology of item responding was discussed. Afterward, the ecological model developed by Zumbo et al. (2015) is introduced as a potential solution for discovering the complex background. Finally, the discussions and proposals for researchers interested in the analysis technics, item developers, and policymakers, especially for ÖSYM, were provided. Based on the inferences, we may conclude that the ecological approach and ecological model offer significant promise for validity studies in the measurement and evaluation of Turkish. Therefore, by Turkey's measurement needs, an ecology-based perspective should be adapted to the measurement of Turkish in the mother tongue. Thus, it will be possible to meet the need for an approach to the measurement and evaluation of the Turkish lesson curriculum.

**Keywords:** Item response process, measurement and evaluation, differential item functioning, reading, ecological model

### 1. Giriş

Türkiye'nin dâhil olduğu ve büyük örneklem grupları ile yürütülen ölçme süreçleri uluslararası ve ulusal seviyede ikiye ayrılabilir (Reyhanlıoğlu ve Tiryaki, 2021). Uluslararası seviyede PISA, TIMSS ve PIRLS gibi araştırmalar ön plana çıkmakta ve Türk eğitim politikalarının belirlenmesi gibi kritik başlıklarda etkili olmaktadır (Güner ve diğerleri, 2014). PISA ve PIRLS sebebiyle de Türkçenin uluslararası platformda ölçülen dil becerilerinin başında okuma gelmektedir (Zuckerman ve diğerleri, 2013). Ana dili olarak Türkçenin ölçülmesi ve değerlendirilmesi ulusal seviyede de Türkiye'nin ölçme ve değerlendirme hedefleri arasında kritik bir noktayı oluşturmaktadır (MEB, 2006, 2018, 2019). İlgili hedefler doğrultusunda ülke genelinde Öğrenci Seçme ve Yerleştirme Merkezi (ÖSYM) ve Milli Eğitim Bakanlığı (MEB) öncülüğünde ölçme süreçleri yürütülmektedir (Akyol ve diğerleri, 2020). Uluslararası seviyedeki uygulamalara paralel olarak Türkiye'deki merkezi imtihanlarda sadece okuma ölçülmektedir.

Her ne kadar okuma becerisi ana dili becerilerinin ölçülmesinde merkezi bir yerde olsa da diğer üç dil becerisi (dinleme, konuşma ve yazma) için çalışmalar sürmektedir. Nitekim PISA, TOEFL, IETLS ve benzeri ölçme sistemlerinden hareketle dört dil becerisinin bilgisayar tabanlı ölçülmesi için MEB (2020) tarafından başlatılan proje devam etmektedir. Pilot çalışma doğrultusunda yayınlanan rapor dışında 2022 yılı içinde 13.000 öğrenciye ek uygulama yapılmıştır (MEB, 2022). Uygulamadan hareketle hem

madde havuzunun genişlemesinin hem de geçerlilik çalışmalarına yönelik odağın artmasının gerekliliği vurgulanmalıdır. Bununla beraber hazırlanan maddelerin geçerlilik analizleri maddenin kendisi ve maddeye verilen tepkilere endeksli olarak yürütülmektedir (Eroğlu ve diğerleri, 2020).

Türkiye için ulusal seviyede yapılan ölçme süreçleri ise liselere geçiş ve üniversitelere geçiş olarak iki başlıkta ele alınabilir. Liselere geçiş imtihanları doğrudan MEB tarafından yürütülürken üniversiteye geçiş imtihanları doğrudan ÖSYM tarafından yürütülmektedir. Uygulanan testler sadece okuma odaklıdır ve yaratma becerisi gibi (Anderson ve diğerleri, 2001) kompleks zihni beceriler ölçülmemektedir. Bu testlerin geliştirilme ve uygulama sürecinin en mühim ayaklarından birini de yine geçerlilik araştırmaları oluşturmaktadır. Paralelde ÖSYM ve MEB tarafından yürütülen ölçme süreçlerine dair düzenli olarak raporlar yayınlanmaktadır (Akyol ve diğerleri, 2020b, 2021; Suna & Özçelik, 2022; Tanberkan ve diğerleri, 2018; Yıldırım ve diğerleri, 2019). Ayrıca bağımsız araştırmacılar tarafından merkezi imtihanlarda kullanılan maddelerin geçerliliğine dair analizler yürütülmektedir (Köroğlu, 2021). Bununla beraber ilgili analizlerde geçerliliğin yine madde ve maddeye verilen tepki boyutunu ifade eden puanlar bazında kaldığı görülmektedir. Merkezi ölçme sistemlerinin maddelerine dair geçerlilik araştırmaları yine maddenin sınırları içinde kalmaktadır. Böylece adeta vakumlanmış bir madde tepki süreciyle katılımcının değerlendirilmek istenen yeteneği başka bir etki yokmuş gibi ele alınmaktadır. Dolayısıyla sınırlı bir bakış açısı ile madde cevaplayıcılarının oluşturduğu heterojen örüntülerin göz ardı edilmesi söz konusudur. Böylece geçerlilik araştırmaları sıklıkla maddenin kendisine odaklanmakta ve katılımcıların ekolojik arka planını (Zumbo ve diğerleri, 2015) hesaba katmamaktadır.

Ölçme araçlarının geliştirilme sürecinde madde cevaplayıcılarının özelliklerine göre oluşan heterojen alt grupların madde tepki sürecine getirdiği çok sayıda faktör unutulmaktadır. İlgili faktörlerle ya da maddenin kendisinden kaynaklı olarak gruplardan biri veya birkaçına haksız avantaj yahut dezavantaj sağlanması hâlinde madde yanlılıkları ortaya çıkabilmektedir (Ackerman, 1992; Mellenbergh, 1989). Zumbo ve Gelin'in (2005) not ettiği gibi bu birincil problem test maddelerinin katılımcılar için adillik ve eşitlik sağlaması bakımından önemlidir. Dolayısıyla madde yanlılıklarının ortaya çıkarılmasında rehber olabilecek bir değer olarak diferansiyel madde fonksiyonu (DMF) seviyesinin belirlenmesi gerekmektedir (Zumbo, 1999). DMF seviyesini belirlemek için kullanılan analizler ölçme araçlarının geliştirilmesinde kullanılabilecek başlıklarda yer bulmaktadır (DeVellis ve Thorpe, 2021). Bununla beraber ekolojik örüntü yakın zamana kadar odak dışı kalmıştır. Türkiye'deki çalışmaların da yakın zamana kadar aynı minvalde olduğu ve cevaplayıcıların getirdiği ekolojik arka planın analiz dışı bırakıldığı görülmektedir (Göktentürk, 2021).

Türkçenin ana dili olarak eğitiminde çok kısır bir gelişme alanına sahip olan diferansiyel madde fonksiyonunun araştırmalarda daha fazla yer bulması ihtiyacı burada vurgulanmalıdır. Zumbo'nun (2007) belirttiği gibi DMF çalışmaları dar bir şekilde yanlılığı tarif etmekle ilgili değildir. Kesinlikle, testte adillik ve eşitlik meseleleri önemlidir ve DMF analizleri doğrudan eğitim politikaları ve mevzuatları tarafından motive edilerek bu konulardan bahsetmektedir. DMF'nin bu şekilde kullanılması bugün hâlâ ehemmiyetini korusa da Zumbo'nun ilk iki nesil DMF araştırmaları testin adilliği ve eşitliğinde konumlanmıştır. DMF de bu sebeple tasarlanmış ve yaratılmıştır. Bununla beraber Zumbo'nun 3. nesil DMF'sinde psikometrik analiz şu amaçlarla da kullanılabilir: 1. Grup mukayeselerinin anlamlı bir şekilde yapılabilmesi için ölçme artefaktlarını ortadan kaldırmak. Bu bazen iç geçerlilik olarak da ifade edilmektedir. 2. Tercüme edilmiş veya uyarlanmış ölçme sistemlerini soruşturmak. 3. Madde tepki süreci ve test performansındaki zihni ve psikososyal yönü anlamaya destek olmak. 4. Farklı gruplardan katılımcılar için bu süreçlerin aynı olup olmadığını araştırmak. Klasik DMF

kullanımının ötesine geçmek kapsayıcı bir teorileştirmeyi gerekli kılmaktadır. Ayrıca kapsayıcı bir teorinin inşası ölçme, değerlendirme ve eğitim için daha büyük etkiler doğurmaktadır. Alan içi akademisyenler tarafından ilgili başlıkta çalışma yapılmamakta genellikle alan dışı araştırmacılar dil becerisi maddelerine çalışmaktadır. Dolayısıyla sınırlı bir çerçeve ve anlayışla sonuçların yorumlanması ihtimali ortaya çıkmaktadır. Hem ekolojik arka planın ihmali hem de diferansiyel madde fonksiyonunun az çalışılmış olması geçerli ölçme araçlarının elde edilmesi yolunda ciddi bir tehdittir. İlgili problem doğrultusunda çalışmada ekolojik arka planın keşfi ile ilgili literatür tartışılmıştır. Sonrasında dünyadaki literatür eleştirel bir gözle incelenmiş ve muhtemel çözümler sıralanmıştır. Ayrıca gelecek araştırmalar ve uygulamalar için politika yapıcılara, madde yazarlarına ve araştırmacılara tavsiye ve teklifler sıralanmıştır. Bu doğrultuda çalışmanın problemleri şu başlıklarda sıralanabilir:

1. Klasik DMF araştırmalarının sınırlılığı ve ana dili olarak Türkçenin ölçülmesi ve değerlendirilmesindeki problemler nelerdir?
2. Ana dili olarak Türkçenin ölçülmesi ve değerlendirilmesinde ekolojik arka plana giden yolda Türkiye ve Dünya'daki teorik arka plan nasıldır?
3. Ekolojik model ve ekoloji tabanlı yaklaşımların ana dili olarak Türkçenin ölçülmesi ve değerlendirilmesindeki muhtemel potansiyeli nedir?

## 2. Klasik DMF araştırmalarının sınırlılığı ve Türkçenin ölçülmesindeki problemler

Ekolojik arka planın diferansiyel madde fonksiyonu (DMF) araştırmalarındaki yerini ifade etmeden evvel DMF araştırmalarının gelişim dönemlerini anlamak yararlıdır. Zumbo (2007) tarafından yapılan tasnife göre birinci nesilde madde yanlılığının belirlenmesi, ikinci nesilde ise yeni istatistik metotlarının keşfi söz konusudur. Detayı bu tartışmanın konusu olmamakla beraber ön plana çıkan DMF belirleme metotları 1. tabloda sıralanmaktadır. Çalışmalar parametrik ve parametrik olmayan veri gruplarına yönelik DMF metotlarına göre sınıflanmıştır (Yang, 2009). Ayrıca metotlar tek biçimli ve tek biçimli olmayan (uniform/non-uniform) DMF'nin belirlenmesi bakımından da ayrılmıştır. Son olarak madde skorlamaları yansıtılmıştır (Alahmadi, 2019). Son hâli ile metotlar şöyledir:

Tablo 1. *Diferansiyel madde fonksiyonunu belirlemek için geliştirilmiş metotlar*

| Metot   | Parametrik/Non-parametrik | Maddelerin Skorlanması (İkili/Çoklu) | Uniform/Non-Uniform   | Kaynak                         |
|---|---------------------------|--------------------------------------|-----------------------|--------------------------------|
| <i>Mantel-Haenszel</i>                                | Non-parametrik            | İ/Ç                                  | Uniform               | Holland ve Thayer (1986)       |
| <i>Breslow-Day</i>                                    | Non-parametrik            | İ/Ç                                  | Non-uniform           | Breslow ve Day (1980)          |
| <i>Standardizasyon</i>                                | Non-parametrik            | İ                                    | Uniform               | Dorans ve diğerleri (1992)     |
| <i>Eş zamanlı madde yanlılığı testi (SIBTEST)</i>     | Non-parametrik            | İ/Ç                                  | Uniform               | Shealy ve Stout (1993)         |
| <i>Dönüştürülmüş madde gücüğü / Angoff'un Deltası</i> | Non-parametrik            | İ/Ç                                  | Uniform               | Angoff ve Ford (1973)          |
| <i>c<sup>2</sup> metodu</i>                           | Parametrik                | İ                                    | Uniform / Non-uniform | Lord (1980); Scheuneman (1979) |

|   |                           |     |                       |   |
|---|---------------------------|-----|-----------------------|---|
| <i>Lojistik regresyon</i>                   | Parametrik/Non-parametrik | i/Ç | Uniform / Non-uniform | Swaminathan ve Rogers (1990)<br>Zumbo (1999)<br>Crane ve diğerleri (2004) |
| <i>Madde tepki kuramı olabilirlik oranı</i> | Parametrik                | i/Ç | Uniform / Non-uniform | Thissen ve diğerleri (1993)   |
| <i>Bonferroni düzeltmesi</i>                | Parametrik                | i/Ç | Uniform / Non-uniform | Stark, Chernyshenko ve Drasgow (2006)                                     |
| <i>Raju'nun alan ölçümü</i>                 | Parametrik                | i   | Uniform / Non-uniform | Raju (1988)   |

Belirtilen metotların maddeye ve maddeye tepki sürecine odaklı olduğu görülmektedir. Her ne kadar bu metotlar geçerliliğe yönelik bazı kavramlar sunsa da temiz bir kavram çerçevesi veya pratik bir süreç tasviri sunmamaktadır (Zumbo, 2017). Metotların mühim bir kısmı *-Mantel-Haenszel, lojistik regresyon, madde tepki kuramı-* eşleştirme grupları üzerinden hareket etse de hepsi istatistiki yaklaşımlarla hareket etmektedir (Zumbo ve diğerleri, 2015); hâlbuki maddeye tepki süreci içinde sadece zihnen değil duygu ve motivasyon yönünden de etkileşme söz konusudur (Leighton ve diğerleri, 2017). Bu süreç ise karmaşık örüntülerden oluşan ekolojik evreninden etkilenebilir. Dolayısıyla sadece ilgili metotlar yoluyla kompleks bir ekolojik arka plandan etkilenen madde tepki süreçlerini tam bir tasvir mümkün değildir.

Madde tepki süreci için oluşturulan sınırlı açıklamaları aşmak adına 20. yüzyılın başında yanlışlık araştırmalarında sosyo-ekolojik, toplum tabanlı ve çevre değişkenlerini göz önünde bulunduran çalışmalara rastlamak mümkündür (Zumbo ve Gelin, 2005). Sistematiik bir şekilde dünyadaki literatür incelendiğinde yakın dönem çalışmalarında ekolojik arka plana daha fazla odaklanıldığı görülmektedir (Li ve diğerleri, 2022). Yine de mevcut istatistiki yaklaşımların DMF araştırmalarını bütün yönüyle görmeye engel olma ihtimali burada akla getirilmelidir. Ekolojik arka planın keşfi esnek bir düşünüş ve jaz icra etmek gibi bir doğaçlamayı gerektirebilmektedir (Maddox ve Zumbo, 2017). Dolayısıyla rutin bakış açılarından uzaklaşmak ve *in vitro* (spesifik şartlar altında inceleme süreci oluşturma) yerine *in vivo* (hayatın içinde ve tabii hâlinde inceleme) bakış açısını geçerlilik araştırmalarında yaygınlaştırmak ihtiyacı oluşmaktadır (Zumbo, 2017).

Türkiye'nin literatüründe ise diferansiyel madde fonksiyonu (DMF) ile ilgili farklı çalışmalarda klasik analizlerin sınırlılığı sürmektedir (Doğan ve Öğretmen, 2010; Göktentürk ve diğerleri, 2021). İlgili sınırlılık için birkaç problemi zikretmek mümkündür: 1. Teorik altyapıdan yoksunluk. 2. Gerçek hayata uygun ölçme (*in vivo*) noktasında yetersizlik. 3. Kapsayıcı bir şekilde ekolojik arka planı tasvirde kısırlık. Dolayısıyla ekolojik arka planın çalışılması için rehber çalışmaların tanınmasının yanında ekolojik arka planı besleyen fikir örüntülerini tanımak faydalıdır. Ayrıca Türkçe eğitiminde yapılan çalışmaların Dünya literatürü ile eş zamanlı olarak artması yararlı olacaktır. Belirtilen sıralama doğrultusunda ilk adımda teorik altyapıyı besleyen paradigmalardan bahsedilmelidir.

### 3. Türkçe için ekolojik arka plana giden yolda teorik arka plan

Türkçe dersi öğretim programında Türkçenin ana dili olarak ölçülmesi ve değerlendirmesi için esnek bir süreç öngörülmektedir (MEB, 2019). Bununla beraber ölçmede faydalanan epistemolojik görüşlerden ve paradigmalardan programda bahsedilmesi söz konusu değildir. Türkçe eğitimi içindeki maddeye

tepki süreci araştırmalarının pozitivist paradigma ekseninde şekillendiği görülmektedir (Göktentürk, 2021). Basit gerçeklik prensibine dayalı bu anlayış (Creswell, 2009) karmaşık bir gerçekliği araştırma bakımından da yetersiz kalmaktadır.

Dünyada maddeye tepki süreçleri ise çok sayıda epistemolojik düşünce örüntüsünden ve sistematik paradigmalardan etkilenmeyi sürdürmektedir (Zumbo, 2017). İlgili alandaki felsefi yaklaşımların alana tanıtılması bir yandan maddeye tepki süreçlerinin ekolojik arka planını inceleyen araştırmaların literatürünü anlamada bir yandan da sosyal bilimler alanında yapılacak farklı çalışmaların beslendiği zemini geliştirmede katkı sunacaktır. Dolayısıyla dünyadaki maddeye tepki sürecine ait yaklaşımları takip ederek Türkçe eğitiminin faydalanabileceği noktalar notlanmalıdır.

Sistematik bir gözle maddeye tepki süreci araştırmaları için Launeanu ve Hubley'in (2017) sunduğu roller burada ufuk açıcudur. Maddeye tepki sürecinin geçerlilik ve test geliştirme araştırmaları için genişleyen rolleri şu başlıklarda ifade edilmiştir: 1. Keşfedicilik ve açıklayıcılık. 2. Yeni odakları analize taşıma. 3. Farklı test süreçleri için deneyli bir yaklaşımı mümkün kılmak. 4. Psikometrik modellemelere katkı sunmak. 5. Süreç tabanlı mikro-metodolojileri uygulamak ve geliştirmek. 6. Tepki sürecini geçerlilik delillerinin diğer kaynakları ile ele almak. 7. Test geliştirme süreçleri için tepki süreçlerini kullanmak. Belirtilen rollerde yapılacak çalışmalar uygulama tabanlı hayatın içinden ölçme çalışmaları (*in vivo*) için potansiyel sunmaktadır.

*In vivo* ölçme sürecinin sağlanması için ekoloji temelli ölçme anlayışına giden yolu besleyen yaklaşımları burada vurgulamak yerindedir. Fox'un (2003) çalışması yanlılık araştırmaları için ekolojik bir yaklaşıma duyulan ihtiyacın üzerinde durmaktadır. McNamara (2007) ise ölçmenin sadece şahsın teste verdiği cevaplarla sınırlı olmadığını, sosyal yönünün de göz önünde bulundurulması gerektiğini vurgulamaktadır. Chalhoub-Deville ise (2003) madde tepki sürecinde katılımcı ve sürece dair özelliklerin kompleks bir örüntüde birbirine bağlı olduğunu ifade etmektedir. Nitekim ekolojik model öncesi çalışmalarda Zumbo ve Gelin (2005) madde tepki sürecinin sosyal yönü ile beraber madde yapısı, toplum ve sürece dair özelliklerini vurgulamıştır. Gerçekliğin karmaşık olarak ele alındığı post-pozitivist dünya görüşü (Creswell, 2009) ile birlikte düşünüldüğünde ekolojik arka planı temele alan bir yaklaşımın geçerlilik araştırmalarındaki rolü daha fonksiyonel bir noktaya erişmektedir.

Zumbo ve meslektaşları bir dizi çalışma ile (Zumbo, 2007, 2017; Zumbo ve diğerleri, 2015; Zumbo ve Gelin, 2005) madde tepki süreci için ekolojik modelin temel prensiplerini klasik madde tepki teorileri (*in vitro*) ile karşılaştırmalı olarak ele almıştır. Zumbo'nun (2007) ifade ettiği gibi araştırmacılar zihnî işlem süreçlerinin daha klasik bir kategorisi olarak düşünülen bilgi işlemeye dayalı zihnî görüşe (cognitive view) odaklanmışlardır. İlgili görüş, testi alan katılımcının mantığına ve akıl yürütmesine, test sürecindeki kontekst ve duygular dikkate alınmaksızın, odaklanmaktadır. Soğuk zihin (cold cognition) olarak da adlandırılan bu süreç bilginin zihinde işlenmesini tarif etmek için duygular ve kontekstten arındırılmayı, ifade eder (Unsworth et al., 2005). Bilinçli veya bilinçsiz bir şekilde örtük veya otomatik olarak yapılan maddeye tepki verme süresi, çalışma hafızası ve işlem hızı bunun içindedir. Bilgi işlemeye yönelik soğuk zihne dayalı bakış açısı tabii olarak ekolojik modelin bilgi işleme süreçlerine aracılık eden kontekste dayalı faktörlerin madde tepki sürecini nasıl etkileyeceği sorusunu doğurmaktadır.

Belirtilen soruya cevap olarak Zumbo'nun ekolojik modelinin anahtar bir özelliğinden bahsedilebilir. Zumbo'nun teorik yaklaşımının ne testi alan katılımcıyı ne de tepki sürecindeki zihnî süreçleri izole edilmiş bir vakumda ele aldığını notlamak mühimdir (Zumbo, 2015). Vakumlanmış bir sürecin aksine

bu teoride testi alanlar tepki sürecine sosyal ve kültürel yönden mevcut özelliklerini ve geçmişlerini de getirirler. Teorisi insanların doğumdan bilitibar kültürü edinmek için geliştiğini ve bir kişinin maruz kaldığı ekolojinin kişinin psikolojisini ve zihni yapısını, madde tepki süreci dâhil, etkilediğini varsayan sosyal bilimlerdeki görüşleri başlangıç alır. Bunu yaparken de test performansını açıklayan değişkenlerin ikili yapısına karşı çalışan ve kontekste dayanan bir açıklama biçimine geçilir. DMF'yi daha geniş bir ölçme geçerliliği meselesine bağlama yolunda ekolojik model "kontekst" meselesini daha ileri bir seviyede parçalar hâlinde birleştirmektedir. Zumbo'nun (2009) görüşünde geçerlilik kontekstleştirilmiş ve pragmatikleştirilmiş bir açıklama olarak ifade edilir, yani çok tabakalı ekolojik arka plan kontekstin kendisidir.

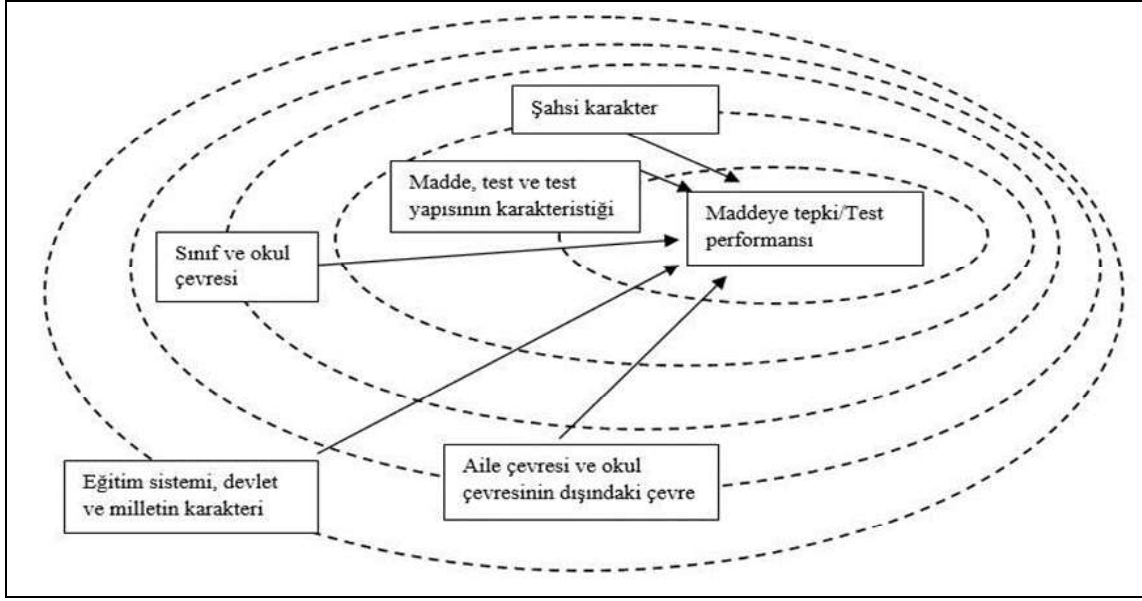
#### 4. Türkçe için ekolojik arka plana giden yolda teorik arka plan

İlgili görüşler doğrultusunda maddeye tepki süreçlerine yönelik ekolojik arka planın araştırılması amacıyla Zumbo ve diğerleri (2015) tarafından ekolojik model teklif edilmiştir. Modeli daha iyi anlamak amacıyla kronolojik bir düzen içinde ekolojik modelin gelişimi aktarılacaktır. İlk olarak DMF'nin sadece problemleri tamir etmek olmadığını hatırlamak gerekmektedir (Zumbo, 1999). Yapılacak bir DMF araştırmasının amaçları ise şu başlıklarda sıralanabilir: 1. Ölçmede eşitlik ve adalet. 2. İç geçerlilik için potansiyel tehditlerle baş etmek. 3. Adapte edilmiş veya tercüme edilmiş ölçme araçları arasında mukayese yapmak. 4. Madde tepki sürecini anlamak. 5. Ölçme değişmezliğine yönelik eksiklikleri araştırmak (Zumbo, 2007). Belirtilen amaçlar doğrultusunda diferansiyel madde fonksiyonu için bu çalışmanın vurguladığı esas nokta maddeye tepki sürecidir. Maddeye tepki süreçlerini anlamak geçerlilik araştırmalarının kilit meselelerinde çözüm ihtimali sunmaktadır. Ayrıca bu süreci anlamak ekolojik arka planı keşfetmede beslenen ana kaynaklardan biridir.

Zumbo'nun eleştirel bakış açısı (2017) aynı zamanda sistematik bir yol haritasının ve kompleks örüntülere şuurlu bir dikkatin yansıması olarak ele alınabilir. Böylece hem ilgili haritayı alana sunmak hem de ekolojik arka plan araştırmalarına bir çağrı oluşturmak için maddeye tepki süreçlerine yönelik örnek bir ekolojik model sunulmuştur (Zumbo ve diğerleri, 2015). Çalışma ekolojik örüntüleri sistematize ederken rehber aldığı Bronfenbrenner'in ekolojik sistem teorisi çocuklar için geliştirilmiştir (Bronfenbrenner, 1979a, 1979b). Modelin teklif edildiği çalışmada ise ilgili teori maddeye tepki süreçlerinin ihtiyaçlarını karşılayacak şekilde uyarlanmıştır. Model Türkçeye aşağıdaki şekilde sunulduğu hâli ile tercüme edilmiştir (Göktentürk, 2021).



Şekil 1. Maddeye tepki süreçlerine yönelik ekolojik model (Zumbo ve diğerlerinden (2015) uyarlanmıştır.)



Model, maddeden hareketle katılımcılara yüklenen vazifelere odaklanmaktadır. Dolayısıyla her madde katılımcının tepki göstereceği bir sürecin başlatıcısı fonksiyonundadır. Süreçte katılımcı farklılıklarından kaynaklanan varyasyonlar devreye girerek tepki sürecini etkilemektedir. Şahsi farklılıkları doğuran varyasyonlar ise maddeler ve katılımcının maddeye yönelik performansına dair sonuçlardan (skorlar) hareketle takip edilebilir.

Okuma veya yazma gibi dil becerileri için test maddeleri sıklıkla kullanılmakta ve dil becerilerinin alt boyutlarını oluşturmaktadır. Dil becerilerinin ölçülmesi ve değerlendirilmesinde ise katılımcıların karakteri ile test karakteri arasında kesişme alanları oluşmaktadır. İlgili kesişme alanları katılımcıların testten negatif veya pozitif etkilenmesi anlamına gelen “testin geri dönüş etkisi” sonucunu (test washback) doğurur. Burada ise ekolojik model yoluyla testin içeriği ve test formatının (kuantum etkisi gibi<sup>4</sup>) eğitimi şekillendirme sürecinde birlikte var olduğu bir süreç temsil edilmektedir. Keza Zumbo’nun da belirttiği gibi (2007) madde yanlılığı çerçevesinde yapılan DMF araştırmaları madde ve maddeden kaynaklı vazifelerle katılımcının karşılaşma sürecini tasvir etmek amacıyla. Böylece modelin bir rehber mahiyetinde olduğu belirtilmelidir. Bu yolla madde ile karşılaşma sürecinin sosyokültürel yönüne de bir pencere açılmaktadır.

Çalışma ekolojik arka plan için vermiş olduğu tabakaları bir teklif olarak sunmaktadır. Dolayısıyla modelin literatürdeki yeri genişledikçe daha farklı tabakalandırmaların görülme ihtimali bulunmaktadır. Ayrıca analiz süreci için kullanılan metotların sadece yollardan biri olarak vurgulanmış olması mühimdir. Dolayısıyla çok sayıda farklı analiz metodu ile arka plandaki ekolojik örüntüye ulaşmak için bir başlangıç ve dünya literatürüne çağrı metni olarak değerlendirilmelidir.

Modelin yayınlanması sonrası kendi literatürünü oluşturmaya başladığı ve farklı değişkenlere yönelik çalışmalara kapı araladığı görülmektedir. Burada ise Zumbo ve Hubley’in (2017) maddeye tepki kuramları çerçevesindeki çalışmalar için daha derinlikli ve kapsayıcılığı yüksek çerçevelere dair çağrısı hatırlanmalıdır. Şu ana kadar yapılan çalışmalarda cinsiyet alt gruplarına yönelik DMF sonuçlarının

4. Klasik mekanikler tarafından ölçilemeyen yapıların kuantum mekaniği ile ölçülebilmesi.

ekolojik arka plandan etkilendiği bulunmuştur (Chen ve Zumbo, 2017). Ayrıca göçmenlik değişkeninin arka planındaki kompleks örüntülerin de etkili olduğu tespit edilmiştir (Roberson ve Zumbo, 2019). Bilhassa okuma ve dil örüntüleri ile ilgili olan muhtemel çalışmalar Türkçenin ölçülmesinde karşılaşılabilecek geçerlilik problemleri için de fırsatlar sunmaktadır.

Türkiye’de modelin kullanılması ile yapılan ilk çalışma Göktentürk (2021) tarafından gerçekleştirilmiştir. Çalışmada PISA okuma maddelerinin kendisinden cevaplayıcının içinde yaşadığı topluma kadar geniş bir değişken evreninden diferansiyel madde fonksiyonu sonuçlarının etkilenmiş olabileceği ortaya konulmuştur. Türkiye’deki merkezi okuma testleri ve dört dil becerisinin ölçüldüğü testler için henüz böyle bir çalışma yapılmamıştır. Dolayısıyla yapılacak çalışmalar bir yandan heterojen bir arka planın muhtemel geçerlilik tehdidini engellemek bir yandan da daha geçerli ölçme araçlarının hazırlanması için yol haritası sunmak bakımından mühimdir.

## 5. Tavsiye ve teklifler

Çalışmamızda yapılan literatür değerlendirmesi Türkiye’deki ölçme sistemleri için yapılacak geçerlilik araştırmalarında ekolojik modelin ciddi bir potansiyele sahip olduğunu ortaya koymaktadır. Bununla beraber ekolojik modelin bir istatistiki metot teklifi sınırlılığında değerlendirilmemesi gerektiği de hatırlanmalıdır (Zumbo ve diğerleri, 2015); zira izah edildiği gibi yeni istatistiki metot madde tepki sürecinde zengin bir ekolojik teori ile iç içedir. Literatürdeki gelişmeler ve Türkçe eğitimi içindeki geçerlilik araştırmalarında diferansiyel madde fonksiyonu çalışmalarının sınırlılığı düşünüldüğünde 3 gruba çağrı yapılması faydalı olacaktır: 1. Politika yapıcılar. 2. Ölçme araçlarını geliştirmede çalışan madde yazarları. 3. Alan içi ve alan dışı araştırmacılar.

Türkçenin eğitimi için yapılan politikalar uygulama gruplarına göre birkaç başlıkta sınıflandırılabilir. Bunların başında ana dili olarak Türkçe öğretimi gelmekte ve Türkçe dersi öğretim programına göre şekillendirilmektedir (MEB, 2019). Programın ölçmede esneklik dışında net bir felsefi arka planı ve ölçme süreci için beslendiği bir yaklaşım söz konusu değildir. Dolayısıyla ölçmenin asli konularından birini oluşturan maddeye tepki süreçleri için ekoloji tabanlı bir yaklaşıma programlarda yer verilmesi gerekmektedir. İlgili yaklaşım ana dili olarak Türkçe öğretimi yanında yabancı dil olarak Türkçe öğretimi ve ikinci dil olarak Türkçe öğretimi için de ciddi bir potansiyel taşımaktadır. Türkiye Maarif Vakfı (2019) tarafından yabancı dil olarak Türkçe öğretimine yönelik hazırlanan programa bakıldığında ölçme ve değerlendirme yaklaşımlarına yer verilmekle beraber ekolojik arka planın yine yer bulmadığı görülmektedir.

Burada ikinci dil olarak Türkçe öğretimine ayrı bir başlığın açılması gerekmektedir. Türkiye Birleşmiş Milletler’in raporlarına göre (McAuliffe & Khadria, 2019) çokluğu Suriye’den olmak üzere Dünya’da en fazla mülteci barındıran ülke konumundadır. Göç sürecindeki mültecilerin mühim bir kısmı için kültürel entegrasyon beklenen bir gelişmedir (Brettell, 2022). Dolayısıyla mevcut mültecilerin yoğun bir şekilde Türk eğitim sistemine katılması beklenen bir sonuçtur. Tabii olarak entegrasyonun sağlanmasının en temel unsuru ise dil becerilerinin ölçülmesi ve değerlendirilmesi olacaktır. Roberson ve Zumbo (2019) tarafından yapılan çalışmada da kısmi olarak açığa çıkarıldığı gibi göçmen grupların dâhil edildiği ölçme süreçlerinde arka plandaki ekolojik yapı mutlaka göz önünde bulundurulmalıdır. Artık milyonları bulan ikinci dil eğitimi alanında da yabancı dil ve ana dili eğitimi gibi Türkiye’nin bir ölçme politikası geliştirmesi kaçınılmazdır. Nihayetinde ise ekolojik örüntüleri dikkate alan bir yaklaşımla ölçme süreçlerini etkileyebilecek yanlışlıkların önüne geçilmesi istenen eğitim hedeflerinin sağlanmasında muhtemel katkılara sahiptir.

İkinci grup olarak madde yazarlarının ekolojik arka planı dikkate aldığı bir tasarlama sürecinin faydası burada vurgulanmalıdır. Her ne kadar merkezi imtihan süreçlerinde bütün heterojen örüntülerin göz önünde bulundurulması imkânsız olsa da ölçme politikalarından doğan amaçlara paralel olarak, göçmen gruplar ve benzeri, muhtemel yanlışlıkların gözetilmesi yerinde olacaktır. Bilhassa okuma testlerinin yaygın olarak kullanıldığı merkezi imtihanlarda (Akyol ve diğerleri, 2020b, 2021; Suna ve Özçelik, 2022; Tanberkan ve diğerleri, 2018; Yıldırım ve diğerleri, 2019) metin seçimi gibi hususlar için bu nokta oldukça mühimdir. Keza merkezi ölçme süreçlerinde ölçülen okuma becerisi için metin seçiminde heterojen arka plan örüntülerinin ihmalı muhtemel yanlışlıklara zemin hazırlayabilmektedir (Gökentürk, 2021). Benzer şekilde literatürde kültürel arka planın metinlerde oluşturduğu yansımalar yoluyla diferansiyel madde fonksiyonu sonuçlarının ekolojik arka planının açıklanmaya çalışıldığı da görülmektedir (Maddox ve diğerleri, 2015). Dolayısıyla madde havuzu oluşturulması sürecinde popülasyon üzerinde geçerlilik tehdidi oluşturabilecek ekolojik örüntülerin belirlenerek madde yazımında göz önünde bulundurulması faydalı olacaktır.

Üçüncü olarak Türkiye'deki araştırmacıların yapacağı muhtemel çalışmalarla ekolojik arka plana dair araştırmalar yapmaları bu çalışmayı hazırlayan araştırmacıların çağrısı arasındadır. Nitekim Akademik Becerilerin İzlenmesi ve Değerlendirilmesi (ABİDE) projesinin maddelerine yönelik yapılan çalışmalar (Elkonca, 2020; Gökentürk ve diğerleri, 2021) ekolojik bir örüntünün araştırılmasına yönelik tavsiyeler sunmaktadır. PISA üzerine elde edilen sonuçlar da bu konuda okuma becerisinin ölçülmesine dair sonuçların ekolojik bir arka plandan etkilendiğini ortaya koymaktadır (Gökentürk, 2021). Bununla beraber merkezi imtihanlarda kullanılan maddelerin detaylı olarak çalışılması gelecekte Türkçenin ölçülmesi ve değerlendirilmesi için daha zengin bir literatüre erişilmesini mümkün kılacaktır.

Türkiye'deki araştırma sürecinin hız kazanması için bir hususu daha burada zikretmek faydalıdır. PISA ve PIRLS gibi okuma becerisinin ölçüldüğü araştırmaların istatistik verileri açık erişimdedir ve araştırmacılar bu konularda analizlerini kolaylıkla tamamlayarak ilgili sistemlerin daha iyi olmasında katkılar sunmaktadır. Türkiye'de MEB ve ÖSYM tarafından yürütülen ölçme süreçlerinde ise araştırmacılar verilere erişmek için uzun süren bir bürokrasiyi aşmak mecburiyetindedir. Aylar sürebilen bu adımların açık erişim ile aşılması alanda ekolojik arka plan çalışmaları için çok daha hızlı ilerleme kaydedilmesini mümkün kılacaktır.

Son olarak politika yapıcılara, madde yazarlarına ve araştırmacılara ekolojinin önemine binaen 3 mesaj sunulmaktadır:

1. Kişinin ekolojik arka planı dikkate alınmadan anlamlı bir şekilde diferansiyel madde fonksiyonu (DMF) çalışmaları yapılamaz.
2. Doğru ve isabetli ölçme ve değerlendirme süreci için katılımcıların ekolojik arka plan Türkiye'nin yapısına doğru şekilde adapte edilmelidir.
3. Dolayısıyla Türkçe Eğitimi'nin anayasası mahiyetindeki Türkçe dersi öğretim programının ölçme yaklaşımına ekolojik perspektif dâhil edilmelidir.

### **Yazarların Katkı Oranı**

Yazarların katkı oranı sırasıyla şu şekildedir: 1. yazar %55, 2. yazar %30, 3. yazar %15.

### ıkar atıřması

Yazarların arasında herhangi bir ıkar atıřması bulunmamaktadır.

### Teřekkr

alıřmaya katkısı sebebiyle TBİTAK 2219-A doktora sonrası arařtırma burslarına teřekkr ederiz.

## New Lens in Assessment and Evaluation of Turkish Language with Zumbo's Critical View of Item Responding

### 1. Introduction

There are two international and national testing programs in which Turkey participates and has a large sample size (Reyhanlıoğlu & Tiryaki, 2021). At the international level, studies such as PISA, TIMSS, and PIRLS come to the fore and become influential in critical topics such as the determination of Turkish education policies (Güner et al., 2014). According to studies such as PISA and PIRLS, reading is a crucial language skill in Turkish education in the international area (Zuckerman et al., 2013). The measurement and evaluation of Turkish as a first language (i.e., mother tongue) constitute a critical point among Turkey's measurement and evaluation objectives at the national level (MEB, 2006, 2018, 2019). Throughout this paper, the terms "mother tongue," "native language," "first language," or the shorthand "L1" will be used interchangeably to refer to the first language learned at home in childhood and is still understood by the person at the time of the assessment or evaluation. In line with the relevant targets, measurement processes under the management of ÖSYM and the Ministry of National Education (MoNE) are carried out throughout the country (Akyol et al., 2020). In similar studies at the international level, only reading is measured in the central exams in Turkey.

Although reading skill is central to measuring L1 skills in Turkey, studies are ongoing for the other three language skills (i.e., listening, speaking, and writing). Likewise, the project initiated by the MoNE (2020) for the computer-based measurement of four language skills, based on PISA, TOEFL, IELTS, and similar measurement systems, continues. The pilot study's report was published, and additional 13,000 students were assessed (Ministry of National Education, 2022). However, the item pool needs to be expanded, and more validity studies are necessary for the developed items. On the other hand, the ecological background was neglected in the validity analysis of the items prepared for computer-based measurement processes (Eroğlu et al., 2020).

The measurement processes at the national level can be classified into two headings: transition to high school and transition to university. While the entrance exams for high schools are carried out directly by the Ministry of National Education (MoNE), the university entrance exams are carried out directly by ÖSYM. The tests focus on reading, and complex cognitive skills such as creativity (Anderson et al., 2001) are not measured. One of the essential pillars of these tests' development and application process is validity studies. In parallel, reports on the measurement processes carried out by ÖSYM and MEB are published regularly (Akyol et al., 2020b, 2021; Suna & Özçelik, 2022; Tanberkan et al., 2018; Yıldırım et al., 2019). Additionally, independent researchers are investigating the validity of the items used in the central exams (Köroğlu, 2021). The validity view, however, is solely based on the results of the item scores; that is, these validity studies remain within the limits of the items, treating them as if the item response is within a vacuum with no other influences than the test taker's ability purported being assessment (e.g., reading ability). Therefore, with a limited perspective, heterogeneous patterns created by item respondents are ignored. Validity studies often focus on the item itself and do not consider the participants' ecological background (e.g., Zumbo, 2017; Zumbo & Gelin, 2005; Zumbo et al., 2015).

In the process of developing measurement tools, a variety of factors result in heterogeneous subgroups. They are formed according to the characteristics of the item responders and the settings (ecology) of

their lives, which are mostly ignored. Item bias may develop if one or more groups receive an unjustified benefit or disadvantage because of associated factors or the item itself (Ackerman, 1992; Mellenbergh, 1989). As Zumbo and Gelin (2005) note, the primary issue is one of consequential matters of test fairness and equity. The differential item function (DIF) level should be determined as a guide for revealing item biases (Zumbo, 1999). The methodologies employed in constructing measurement tools include the analysis performed to calculate the DIF level (DeVellis ve Thorpe, 2021). However, the ecological background has found no place in these methods. Until recently, studies in Turkey were conducted similarly, and the respondents' ecological background was not considered (Göktentürk, 2021).

The importance of DIF studies, which has a very limited development area in Turkish education, should be emphasized here. As Zumbo (2007) highlights, DIF studies are not narrowly about identifying bias. Certainly, matters of fairness and equity in testing are important, and DIF analyses speak to this directly motivated by policy and legislation. Although this use of DIF is still important today, this is where Zumbo's first two generations of DIF were clearly situated and DIF was conceived of and created with this purpose in mind. However, in Zumbo's third-generation-DIF this type of psychometric analysis can also be used to: (a) rule out measurement artifacts so that one can make group comparisons meaningfully, sometimes referred to as internal validity, (b) investigate the comparability of translated and/or adapted measures, (c) help understand the cognitive and/or psychosocial processes of item responding and test performance, and (d) investigating whether these processes are the same for different groups of individuals. Moving beyond the traditional DIF use in the context of test bias has demanded developments for expansive DIF theorizing, but the expansive view has greater impact for assessment, evaluation, and education. In-field academicians do not work on the topic; non-field researchers generally work on language skills. Therefore, the possibility of interpreting the results with a limited framework and understanding arises. The understudied DIF and the disregard for the ecological setting should be considered critical issues. In line with the problem, the discovery of the ecological background was highlighted as a crucial issue of validity in this study. Afterward, possible solutions are listed through a critical evaluation of the literature. In addition, recommendations to policymakers, item developers, and researchers are listed. In this direction, the problems of the study can be listed under the following:

1. What are the limitations of conventional DIF studies and the problems in measuring and evaluating Turkish as a mother tongue?
2. What is the theoretical background in Turkey and in the world on the way to ecological background in the measurement and evaluation of Turkish as a mother tongue?
3. What is the potential potential of ecological models and ecology-based approaches in measuring and evaluating Turkish as a mother tongue?

## 2. Limitations of Conventional DIF Studies and Problems in Measuring Turkish

Before describing the function of ecological background in DIF investigations, it is helpful to understand the stages of DIF study development. According to the classification made by Zumbo (2007), determining item bias was the focal point in the first-generation DIF studies. In the second generation, new statistical methods were discovered. Although the details are not the subject of this discussion, the most widely used DIF analysis methods that come to the fore are listed in Table 1. Studies were classified according to parametric and non-parametric DIF methods for pre-existing data groups (Yang, 2009). In

addition, the methods are separated in terms of determining uniform and non-uniform DIF. Finally, item scores are presented (Alahmadi, 2019). The final methods are as follows:

Table 1. *Methods to investigate DIF*

| Method   | Parametric/Non-parametric | Item scores<br>Dichotomous/<br>Polytomous | Uniform/Non-Uniform   | Reference  |
|--|---------------------------|---|-----------------------|--|
| <i>Mantel-Haenszel</i>                                     | Non-parametric            | D/P                                       | Uniform               | Holland & Thayer (1986)  |
| <i>Breslow-Day</i>   | Non-parametric            | D/P                                       | Non-uniform           | Breslow & Day (1980)   |
| <i>Standardizasyon</i>                                     | Non-parametric            | D   | Uniform               | Dorans et al. (1992)   |
| <i>Simultaneous Item Bias Test (SIBTEST)</i>               | Non-parametric            | D/P                                       | Uniform               | Shealy & Stout (1993)  |
| <i>Transformed item difficulties / Angoff's Delta Plot</i> | Non-parametric            | D/P                                       | Uniform               | Angoff & Ford (1973)   |
| <i>c<sup>2</sup> method</i>                                | Parametric                | D   | Uniform / Non-uniform | Lord (1980); Scheuneman (1979)                                     |
| <i>Logistic regression</i>                                 | Parametrik/Non-parametric | D/P                                       | Uniform / Non-uniform | Swaminathan & Rogers (1990)<br>Zumbo (1999)<br>Crane et al. (2004) |
| <i>Item response theory likelihood ratio</i>               | Parametric                | D/P                                       | Uniform / Non-uniform | Thissen et al. (1993)  |
| <i>Bonferroni correction</i>                               | Parametric                | D/P                                       | Uniform / Non-uniform | Stark, Chernyshenko & Drasgow (2006)                               |
| <i>Raju's area measures</i>                                | Parametric                | D   | Uniform / Non-uniform | Raju (1988)  |

The methods in Table 1 focus on the item and the item response process. As Zumbo (2007) shows, these techniques can offer some insights into test validity. However, they do not offer a precise framework or a useful explanation of the process (Zumbo, 2017). Although many methods - *Mantel-Haenszel*, *logistic regression*, and *item response theory* - use matching groups, they all act as statistical approaches (Zumbo et al., 2015). However, in the item response process, the encounter is not only mental but also about emotion and motivation (Leighton et al., 2017). On the other hand, the ecological background, which comprises intricate patterns, has the potential to influence item responses. As a result, relying solely on these methodologies is insufficient to fully describe the item response mechanisms influenced by a complex ecological context.

It was possible to locate studies that considered socio-ecological, society-based, and environmental variables in bias research at the beginning of the 20th century to meet the restricted explanations for the item response process (Zumbo & Gelin, 2005). When the world literature is examined systematically, the ecological background is more focused in recent studies (Li et al., 2022). However, we should remember that current statistical approaches may prevent us from seeing all aspects of DIF research. Exploring the ecological background may require flexible thinking and improvisation, such as

experienced when playing jazz (Maddox & Zumbo, 2017). Therefore, there is a need to move away from routine perspectives and to popularize the *in vivo* (i.e., examination in life and natural state) perspective in validity studies instead of *in vitro* (i.e., creating an examination process under specific conditions) (Zumbo, 2017).

In the literature of Turkey, the limitations of conventional analyzes continue in several studies on DIF (Doğan & Öğretmen, 2010; Göktentürk et al., 2021). It seems possible to mention a few problems with the relevant limitation: 1. Lack of theoretical background 2. Inadequate real-life measurement point (*in vivo*) 3. Barrenness in depicting the ecological background holistically. Consequently, combining Turkish education with global literature studies would be advantageous. Considering this, the ecological viewpoint's theoretical foundation should be addressed as the first step.

### 3. Theoretical Background on the Path of Ecological Background of Turkish

The Turkish lesson curriculum envisages a flexible process to measure and evaluate Turkish (Ministry of National Education, 2019). However, it is out of the question to talk about epistemological views and paradigms used in the measurement philosophy of the curriculum. The studies on the item response process in Turkish education are shaped along the axis of the positivist paradigm (Göktentürk, 2021). This understanding, based on the simple reality (Creswell, 2009), is also insufficient for investigating a complex reality.

Specifically, the item response process continues to be influenced by many epistemological views and paradigms (Zumbo, 2017). Introducing the philosophical approaches in the field will contribute to the understanding of the literature examining the ecological background of the item response processes and develop the basis on which different studies can be carried out in the field of social sciences. Therefore, by following the approaches to the item reaction process in the world, the points that Turkish education can benefit from should be noted.

The roles presented by Launeanu and Hubley (2017) for the studies on the item response process are eye-opening and particularly important. These include exploratory and explanatory roles; changing the unit of analysis; using the response process in experimental approaches to validation; psychometric modeling of response processes; development and application of process-based micro-methodologies; using response processes with other sources of validity evidence; and using response processes in the test development process. Studies in the specified roles offer the potential for application-based *in vivo* measurement studies.

It is appropriate to highlight the approaches that led the way to an ecology-based measurement approach to enable the *in vivo* measurement process. Fox's (2003) study emphasizes the need for an ecological approach to item bias research. McNamara (2007), on the other hand, emphasizes that the measurement is not limited to the answers given by the individual to the test. However, the social aspect should also be taken into consideration. Further, Chalhoub-Deville (2003) states that participant and procedural features are interconnected in a complex pattern in the item response process. Along with their studies of ecological models, Zumbo and Gelin (2005) emphasized the social aspect of the item response process and the characteristics of item structure, society, and process. When considered together with the post-positivist worldview (Creswell, 2009), in which reality is assumed to be complex, the role of an approach based on an ecological background in validity studies reaches a more functional point.



In a series of papers, Zumbo and his colleagues (e.g., Zumbo, 2007; Zumbo, 2017; Zumbo & Gelin, 2005; Zumbo et al., 2015) addressed several of the central tenets of the ecological model of item responding by contrasting it to conventional (in vitro) views of item responding. As Zumbo (2007) notes, in the first two generations of DIF research, when researchers pursued the question of why an item was displaying DIF, they focused on an information-processing cognitive view typically thought of as the more classic category of cognition, focusing on the test taker's basic logic and reason without concern for their emotional or social context- cold cognition is a concept used to describe information processing in its strictest applications stripped of emotions or context – be it conscious or unconscious (Unsworth et al., 2005), that is, implicit or automatic such as response time, working memory and processing speed. From this information-processing cold cognition point of view, the question naturally arises of how the ecological model's contextual factors can could affect item responses by mediating the cognitive processes normally assumed to generate item responses

In response to this question, one can elucidate a key feature that sets Zumbo's ecological model of item responding from others. It is important to note that Zumbo's theorizing adheres to the view that neither the test taker nor the cognitive processes in item responding are isolated in a vacuum (Zumbo, 2015). Instead, test takers bring their social and cultural present and history to test taking. His theorizing takes as its starting point the widely received view in the broader social sciences that human beings have evolved to acquire culture from birth and that the culture to which an individual is exposed, and the ecology of their lives, affects their basic psychology and cognition, including, in our case, item responding. In so doing, one moves to a contextualized form of explanation that works against a binary structure of variables that explain test performance. In linking DIF to the broader issue of measurement validity, the ecological model further articulates what "context" means in Zumbo's (2009) view of validity as contextualized and pragmatic explanation—that is, the multilayered ecology is the context.

#### **4. The Ecological Model: A Potential Solution for Examining the Ecological Background of the Item Response Process**

An ecological model has been proposed by Zumbo et al. (2015) in order to investigate the ecological background of item-response processes in line with related approaches. To better understand the model, the development of the ecological model will be presented in chronological order. First, it is important to remember that DIF is not just about repairing problematic items (Zumbo, 1999). The objectives of DIF research can be listed under the following headings: (1) fairness and equity, (2) dealing with a possible threat to internal validity, (3) investigating the comparability of translated or adapted measures, (4) trying to understand item response process, and (5) investigating lack of invariance (Zumbo, 2007). In line with the stated purposes, the main point emphasized by this study for DIF is the item response process. Understanding the item response process offers potential solutions to key issues of validity studies. Also, understanding this process is one of the main resources to explore the ecological background.

Zumbo's critical perspective (2017) can also be considered a reflection of a systematic roadmap and conscious attention to complex patterns. Thus, Zumbo et al. (2015) presented a prototypical ecological model for item response processes to present the relevant map to the field and to create a call for ecological background research. Bronfenbrenner's ecological system theory, which the study took as a guide while systematizing ecological patterns, emerged out of the field of human development (Bronfenbrenner, 1979a, 1979b). In the study in which the model was proposed, the theory

was adapted to meet the needs of the item response process. The model has been translated into Turkish, as presented below (Göktentürk, 2021).

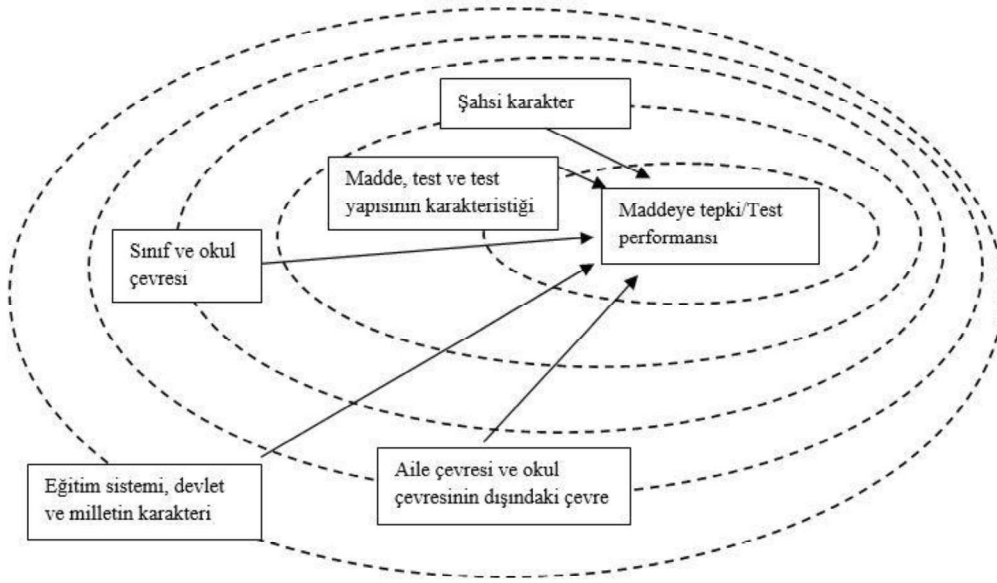


Figure 1. An ecological model of item responding (adapted from Zumbo et al. (2015))

The model focuses on the tasks assigned to the participants based on the item. Therefore, each item functions as the initiator of a process which the participant will enact. Variations arising from individual differences in the item response process come into play and affect the response process. Variations that cause individual differences can be followed by looking at the results of the items and the participant's performance on the item.

Test items for language skills such as reading or writing are frequently used and constitute the sub-dimensions of language skills. In the measurement and evaluation of language skills, there are intersections between the character of the participants and the test character. Sometimes this is called "test washback," which represents, for example, how test content and format can co-exist (like a quantum effect) to constrain teaching. Likewise, as stated by Zumbo (2007), DIF studies conducted within the framework of item bias aim to describe the process of encountering the participant with the item and item-related tasks. Thus, the ecological model's layers should be considered as a guide. In this way, a window is opened to the socio-cultural aspect of the process of when test takers encounter (interact with) items.

The study presents the layers of the ecological model as one possible prototype. Therefore, seeing more or different ecological layers as the literature expands is possible. It is crucial to note that the techniques employed for analysis are just one method among several. Therefore, the ecological item response model should be considered a starting point and a call to global literature to discover the ecological background.

The publication of the model started to create its own literature and opened the door to studies on different variables. On the other hand, Zumbo and Hubley's (2017) call for deeper and more inclusive frameworks should be kept in mind. Studies conducted so far have found that DIF results for gender subgroups are affected by ecological background (Chen & Zumbo, 2017). Furthermore, a discovery about

social belonging as a part of complex patterns in the backgrounds of immigrants has been made (Roberson & Zumbo, 2019). Possible studies on reading and language patterns also offer opportunities to measure Turkish.

The first study using the model in Turkey was carried out by Gökentürk (2021). The study showed that a broad ecological background might impact the DIF results of PISA reading items (i.e., from the items to the society in which the respondent lives). Such a study has not yet been conducted for central reading tests in Turkey. Therefore, the studies to be carried out are important in preventing a possible threat to the validity of a heterogeneous background and providing a roadmap for the preparation of more valid measurement tools.

## 5. Recommendations

This review reveals that the ecological model has substantial potential in validity studies for measurement systems in Turkey. However, it is important to note that the ecological model should not be considered solely as a proposal of a novel statistical method (Zumbo et al., 2015) because, as we have shown, the novel statistical method is embedded in a novel rich ecological theory of item responding. From the vantage point of this ecological model, DIF validity studies have had limitations in the Turkish education literature. Therefore, we call on policymakers, test developers, item writers, and in-field and non-field researchers to begin to consider Zumbo's richer ecological model of item responding in their praxis.

Turkish education policies can be classified under various headings according to application groups. Turkish teaching as L1 (i.e., mother tongue) is one of them, and the Turkish course is shaped according to the curriculum (Ministry of National Education, 2019). There is no philosophical approach in the program's measurement policy except openness to flexibility. Therefore, one of the essential measurement issues should be included in the program for an ecology-based approach to test design, item development, and test validation that centers on the ecology of the response process. This approach has a significant potential for teaching Turkish as a foreign language, an L1, and a second language to discover the complex ecological background. Here, we can note that the assessment methodologies are included in the program created by the Turkish Maarif Foundation (2019) for teaching Turkish as a foreign language, but the ecological context does not reappear.

A separate title should be created for teaching Turkish as a second language. According to the reports of the United Nations (McAuliffe & Khadria, 2019), Turkey is the country hosting the highest number of refugees in the world, mostly from Syria. Cultural integration is an expected development for many refugees in the migration process (Brettell, 2022). As a result, it makes sense to assume that current refugees will actively participate in the Turkish educational system. The assessment of language skills will undoubtedly be a fundamental component of integration. It is inevitable for Turkey to develop a measurement policy in the field of second language education, which now counts for millions, just like a foreign language and L1 education. Finally, with an approach that considers ecological background, it is possible to contribute to accomplishing the desired educational goals while preventing biases that may affect the measurement process.

Secondly, we should also emphasize the benefit of designing a process where the test and item developers consider the ecological background. Although it is impossible to consider all heterogeneous patterns in central examination processes, it would be appropriate to consider possible biases, such as

immigrant groups, in parallel with the objectives of measurement policies. This is critical for issues such as text selection, particularly in central exams where reading tests are commonly used (Akyol et al., 2020b, 2021; Suna & Özçelik, 2022; Tanberkan et al., 2018; Yıldırım et al., 2019). Likewise, neglecting heterogeneous backgrounds in selecting the text of the reading tests in the central assessment can cause possible biases (Göktentürk, 2021). Similarly, the ecological background of the results of DIF can be explained through the reflections on the cultural background in the texts (Maddox et al., 2015). As a result, it would be advantageous to identify ecological trends that could jeopardize the validity and account for them when writing the items in the item pool.

Third, one of the calls made by the researchers who prepared this study is for researchers in Turkey to conduct research on the ecological background with potential studies that reflect the ecology of Turkey's social and educational systems. As a matter of fact, studies on the items of the Monitoring and Evaluation of Academic Skills (ABİDE) project (Elkonca, 2020; Göktentürk et al., 2021) offer recommendations for investigating an ecological pattern. The results of PISA also reveal that the DIF results of reading items are affected by an ecological background (Göktentürk, 2021). However, a detailed study of the items used in the central exams will make it possible to access richer literature to measure and evaluate Turkish in the future.

Finally, it is useful to mention one more point to speed up the studies in Turkey. Researchers can easily complete their analyses on these issues and contribute to improving the relevant systems because the statistical data of studies such as PISA and PIRLS, in which reading skills are measured, is open access. Researchers must overcome a long bureaucracy to access data in the measurement processes carried out by the Ministry of National Education (MoNE) and ÖSYM in Turkey. Overcoming these steps, which may take months, with open access, will make it possible to make much faster progress for ecological background studies in the field.

Finally, three messages are presented to policymakers, item developers, and researchers regarding the importance of ecology:

1. It is impossible to conduct differential item function (DIF) investigations meaningfully without considering the subject's ecological background.
2. The ecological background should be tailored to the Turkish setting for proper and accurate measurement and evaluation.
3. Therefore, the ecological perspective should be included in the measurement approach of the Turkish lesson curriculum, which is the constitution of Turkish Education.

#### **Authors contribution rate**

The authors contributed to the study as the following: 1. 55%. 2. 30%. 3. 15%.

#### **Conflict of interest**

The authors declare no competing interests.

Thanks

We thank TÜBİTAK 2219-A post-doctoral research fellowships for their contribution to the study.

### Kaynakça

- Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement*, 29(1), 67–91. <https://doi.org/10.1111/j.1745-3984.1992.tb00368.x>
- Akyol, İ., Özban, A. Y., Öçal, A. A., Kişioğlu, Y., Barış Pekmezci, F., Yılmaz Koğar, E., Kodallı Sevindir, H., Ünal, C., & Sevindir, M. (2021). *2021-YKS değerlendirme raporu* (Değerlendirme Raporu No. 27; Değerlendirme Raporları Serisi, p. 111). ÖSYM. <https://dokuman.osym.gov.tr/pdfdokuman/2021/GENEL/yksdegrapor24122021.pdf>
- Akyol, Özban, A. Y., Öçal, A. A., Kişioğlu, Y., Kodallı Sevindir, H., Ünal, C., Yılmaz Koğar, E., Sevindir, M., Dinarcı, S., Okumuş, A. Ö., Yörük, Y., Bozdağ Kasap, A., Çelik, D., Pehlivanoglu, B., & Kaplan, E. (2020a). *2020 YKS-Değerlendirme Raporu* (Değerlendirme Raporu No. 17; Değerlendirme Raporları Serisi, Issue 17, p. 115). ÖSYM.
- Akyol, Özban, A. Y., Öçal, A. A., Kişioğlu, Y., Kodallı Sevindir, H., Ünal, C., Yılmaz Koğar, E., Sevindir, M., Dinarcı, S., Okumuş, A. Ö., Yörük, Y., Bozdağ Kasap, A., Çelik, D., Pehlivanoglu, B., & Kaplan, E. (2020b). *2020-YKS değerlendirme raporu* (Değerlendirme Raporu No. 17; Değerlendirme Raporları Serisi, p. 115). ÖSYM.
- Alahmadi, M. T. S. (2019). *Evaluating possible bias in a giftedness assessment: An empirical comparison of IRT and non-IRT (parametric and non-parametric) techniques for detecting differential item functioning* [Doctoral Dissertation, Morgan State University]. <https://mdsoar.org/handle/11603/17669>
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Angoff, W. H., & Ford, S. F. (1973). Item-race interaction on a test of scholastic aptitude. *Journal of Educational Measurement*, 10(2), 95–105. <https://doi.org/10.1111/j.1745-3984.1973.tb00787.x>
- Breslow, N. E., & Day, N. (1980). *Statistical methods in cancer research Volume 1—The analysis of case-control studies*. International Agency for Research on Cancer.
- Brettell, C. B. (2022). Theorizing migration in Anthropology: The cultural, social, phenomenological, and embodied dimensions of human mobility. In C. B. Brettell & J. F. Hollifield (Eds.), *Migration theory: Talking across* (4th ed., pp. 194–231). Routledge.
- Bronfenbrenner, U. (1979a). Contexts of child rearing: Problems and prospects. *American Psychologist*, 34(10), 844–850. <https://doi.org/10.1037/0003-066X.34.10.844>
- Bronfenbrenner, U. (1979b). *The ecology of human development: Experiments by nature and design*. Harvard University Press.
- Chalhoub-Deville, M. (2003). Second language interaction: Current perspectives and future trends. *Language Testing*, 20(4), 369–383. <https://doi.org/10.1191/0265532203lt2640a>
- Chen, M. Y., & Zumbo, B. D. (2017). Ecological framework of item responding as validity evidence: An application of multilevel DIF modeling using PISA data. In *Understanding and investigating response processes in validation research* (pp. 53–68). Springer.
- Crane, P. K., Belle, G. van, & Larson, E. B. (2004). Test bias in a cognitive test: Differential item functioning in the CASI. *Statistics in Medicine*, 23(2), 241–256. <https://doi.org/10.1002/sim.1713>
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Sage Publications.
- DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). Sage Publications.

- Doğan, N., & Öğretmen, T. (2010). Değişen madde fonksiyonunu belirlemede mantel-haenszel, ki-kare ve lojistik regresyon tekniklerinin karşılaştırılması. *Eğitim ve Bilim*, 33(148), 100–112.
- Dorans, N. J., Schmitt, A. P., & Bleistein, C. A. (1992). The standardization approach to assessing comprehensive differential item functioning. *Journal of Educational Measurement*, 29(4), 309–319. <https://doi.org/10.1111/j.1745-3984.1992.tb00379.x>
- Elkonca, F. (2020). *ABİDE özyeterlilik ölçeği DMF kaynaklarının gizil sınıf yaklaşımıyla incelenmesi* [Doktora Tezi]. Gazi Üniversitesi.
- Eroğlu, E., Canidemir, A., Suna, H. E., Tanberkan, H., & Altun, Ü. (2020). *Dört beceride Türkçe dil sınavı: Pilot çalışma sonuçları* (Eğitim Analiz ve Değerlendirme Raporları Serisi No. 11; p. 54). Millî Eğitim Bakanlığı. [https://www.meb.gov.tr/meb\\_iys\\_dosyalar/2020\\_01/20094146\\_Dort\\_Beceride\\_Turkce\\_Dil\\_Sinavi\\_Ocak\\_2020.pdf](https://www.meb.gov.tr/meb_iys_dosyalar/2020_01/20094146_Dort_Beceride_Turkce_Dil_Sinavi_Ocak_2020.pdf)
- Fox, J. D. (2003). From products to process: An ecological approach to bias detection. *International Journal of Testing*, 3(1), 21–47. [https://doi.org/10.1207/S15327574IJTO301\\_2](https://doi.org/10.1207/S15327574IJTO301_2)
- Göktentürk, T. (2021). *PISA okuma testinin ekolojik modele göre yanlılık göre yanlılık analizi: Türkiye örneği* [Doktora Tezi]. Yıldız Teknik Üniversitesi.
- Göktentürk, T., DemİR, İ., & Arici, A. F. (2021). PISA'nın ışığında geliştirilen ABİDE projesinde okuma bakımından ne hâldeyiz? Bir geçerlilik çalışması. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi*, 22, 657–665. <https://doi.org/10.29000/rumelide.897081>
- Güner, H., Çelebi, N., Kaya, G. T., & Korumaz, M. (2014). Neoliberal eğitim politikaları ve eğitimde fırsat eşitliği bağlamında uluslararası sınavların (PISA, TIMSS ve PIRLS) analizi. *Journal of History Culture and Art Research*, 3(3), 33–75.
- Holland, P. W., & Thayer, D. T. (1986). Differential item functioning and the Mantel-Haenszel procedure. *ETS Research Report Series*, 1986(2), i–24. <https://doi.org/10.1002/j.2330-8516.1986.tb00186.x>
- Köroğlu, M. (2021). *The validity study of central exam scores used for placement in secondary education institutions* [Doctoral Dissertation, Hacettepe University]. <http://www.openaccess.hacettepe.edu.tr:8080/xmlui/bitstream/handle/11655/24919/%c4%boMZASIZ-MUSTAFA%20K%c3%96RO%c4%9eLU%20DR.%20TEZ%20-%20Kopya.pdf?sequence=1&isAllowed=y>
- Launeanu, M., & Hubley, A. M. (2017). Some observations on response processes research and its future theoretical and methodological directions. In B. D. Zumbo & A. Hubley (Eds.), *Understanding and investigating response processes in validation research* (p. XIII+370). Springer.
- Leighton, J. P., Tang, W., & Guo, Q. (2017). Response processes and validity evidence: Controlling for emotions in think aloud interviews. In B. D. Zumbo & A. Hubley (Eds.), *Understanding and investigating response processes in validation research* (p. XIII+370). Springer.
- Li, H., Hunter, C. V., & Bialo, J. A. (2022). A revisit of Zumbo's third generation DIF: How are we doing in language testing? *Language Assessment Quarterly*, 19(1), 27–53. <https://doi.org/10.1080/15434303.2021.1963253>
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge.
- Maddox, B., & Zumbo, B. D. (2017). Observing testing situations: Validation as Jazz. In B. D. Zumbo & A. M. Hubley (Eds.), *Understanding and investigating response processes in validation research*. Springer.
- Maddox, B., Zumbo, B. D., Tay-Lim, B., & Qu, D. (Iris). (2015). An anthropologist among the psychometricians: Assessment events, ethnography, and differential item functioning in the Mongolian Gobi. *International Journal of Testing*, 15(4), 291–309. <https://doi.org/10.1080/15305058.2015.1017103>
- McAuliffe, M., & Khadria, B. (Eds.). (2019). *World Migration Report 2020*. International Organization for Migration (IOM).

- McNamara, T. (2007). Language testing: A question of context. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. E. Turner, & C. Doe (Eds.), *Language Testing Reconsidered* (pp. 131–138). University of Ottawa Press. <https://doi.org/10.2307/j.ctt1ckpccf>
- MEB. (2006). *İlköğretim Türkçe dersi (6, 7, 8. Sınıflar) öğretim programı*. Millî Eğitim Bakanlığı.
- MEB. (2018). *Ortaöğretim Türk dili ve edebiyatı dersi (9, 10, 11 ve 12. Sınıflar) öğretim programı*. Millî Eğitim Bakanlığı Yayınları. <http://mufredat.meb.gov.tr/Dosyalar/201812103715395-T%C3%9CRKD%C4%B0L%C4%B0EDEB%C4%B0YAT%20PROGRAM.pdf>
- MEB. (2019). *Türkçe dersi öğretim programı (ilkokul ve ortaokul 1, 2, 3, 4, 5, 6, 7 ve 8. Sınıflar)*. Milli Eğitim Bakanlığı.
- MEB. (2022, November 27). *MEB'den dört temel beceride Türkçe dil sınavı*. T.C. Millî Eğitim Bakanlığı Haberler. <https://www.meb.gov.tr/mebden-dort-temel-beceride-turkce-dil-sinavi/haber/26629/tr>
- Mellenbergh, G. J. (1989). Item Bias and Item Response Theory. *International Journal of Educational Research*, 13(2), 127–143. [https://doi.org/10.1016/0883-0355\(89\)90002-5](https://doi.org/10.1016/0883-0355(89)90002-5)
- Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika*, 53(4), 495–502. <https://doi.org/10.1007/BF02294403>
- Reyhanhoğlu, Ç., & Tiryaki, İ. (2021). Ülkemizde gerçekleştirilen ölçme ve değerlendirme faaliyetlerine genel bir bakış. *Uluslararası Türk Eğitim Bilimleri Dergisi*, 9(16), 70–93. <https://doi.org/10.46778/goputeb.766689>
- Roberson, N. D., & Zumbo, B. D. (2019). Migration background in PISA's measure of social belonging: Using a diffractive lens to interpret multi-method DIF studies. *International Journal of Testing*, 19(4), 363–389.
- Scheuneman, J. (1979). A method of assessing bias in test items. *Journal of Educational Measurement*, 16(3), 143–152.
- Shealy, R., & Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika*, 58(2), 159–194. <https://doi.org/10.1007/BF02294572>
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology*, 91(6), 1292–1306. <https://doi.org/10.1037/0021-9010.91.6.1292>
- Suna, H. E., & Özçelik, H. (2022). *2022 ortaöğretim kurumlarına ilişkin merkezi sınav*. Millî Eğitim Bakanlığı.
- Swaminathan, H., & Rogers, H. J. (1990). Detecting Differential Item Functioning Using Logistic Regression Procedures. *Journal of Educational Measurement*, 27(4), 361–370. <https://doi.org/10.1111/j.1745-3984.1990.tb00754.x>
- Tanberkan, H., Suna, H. E., & Eroğlu, E. (2018). *2018 YKS değerlendirme raporu* (Değerlendirme Raporu No. 9; Değerlendirme Raporları Serisi, p. 82). ÖSYM. <https://dokuman.osym.gov.tr/pdfdokuman/2018/GENEL/YKSDeRAPOR06082018.pdf>
- Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 67–114). Routledge.
- Türkiye Maarif Vakfı. (2019). *Yabancı dil olarak Türkçe öğretimi programı*. Türkiye Maarif Vakfı.
- Unsworth, N., Heitz, R. P., & Engle, R. W. (2005). Working Memory Capacity in Hot and Cold Cognition. In R. W. Engle, G. Sedek, U. von Hecker, & D. N. McIntosh (Eds.), *Cognitive Limitations in Aging and Psychopathology* (1st ed., pp. 19–43). Cambridge University Press. <https://doi.org/10.1017/CBO9780511720413.003>

- Yang, C. (2009). *Development and application of the logistic regression with factor scores (LRFS) method in differential item functioning (DIF) detection for dichotomized variables* [Doctoral Dissertation]. Medical University of South Carolina.
- Yıldırım, S., Okumuş, A. Ö., & Yörük, Y. (2019). *2019 YKS değerlendirme raporu* (Değerlendirme Raporu No. 15; Değerlendirme Raporları Serisi, p. 129). ÖSYM.  
<https://dokuman.osym.gov.tr/pdfdokuman/2019/GENEL/yksDegRaporweb03092019.pdf>
- Zuckerman, G. A., Kovaleva, G. S., & Kuznetsova, M. I. (2013). Between PIRLS and PISA: The advancement of reading literacy in a 10–15-year-old cohort. *Learning and Individual Differences*, 26, 64–73. <https://doi.org/10.1016/j.lindif.2013.05.001>
- Zumbo, B. D. (1999a). A Handbook on the Theory and Methods of Differential Item Functioning (DIF). *Ottawa: National Defense Headquarters*, 1–57.
- Zumbo, B. D. (1999b). A handbook on the theory and methods of differential item functioning (DIF). *Ottawa: National Defense Headquarters*, 1–57.
- Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, 4(2), 223–233.  
<https://doi.org/10.1080/15434300701375832>
- Zumbo, B. D. (2009). Validity as contextualized and pragmatic explanation, and its implications for validation practice. In R. W. Lissitz (Ed.), *The Concept of Validity: Revisions, New Directions and Applications* (pp. 65–82). IAP Information Age Publishing.
- Zumbo, B. D. (2015). *Consequences, side effects and the ecology of testing: Keys to considering assessment “In Vivo.”* Keynote address, the annual meeting of the Association for Educational Assessment - Europe (AEA-Europe), Glasgow, Scotland.  
[https://www.youtube.com/watch?v=oL6Lr2BzuSQ&ab\\_channel=BrunoD.Zumbo](https://www.youtube.com/watch?v=oL6Lr2BzuSQ&ab_channel=BrunoD.Zumbo)
- Zumbo, B. D. (2017). Trending away from routine procedures, toward an ecologically informed In Vivo view of validation practices. *Measurement: Interdisciplinary Research and Perspectives*, 15(3–4), 137–139.
- Zumbo, B. D., & Gelin, M. N. (2005). A matter of test bias in educational policy research: Bringing the context into picture by investigating sociological/community moderated (or mediated) test and item bias. *Journal of Educational Research & Policy Studies*, 5(1), 1–23.
- Zumbo, B. D., & Hubley, A. M. (2017). *Understanding and investigating response processes in validation research* (Vol. 26). Springer.
- Zumbo, B. D., Liu, Y., Wu, A. D., Shear, B. R., Olvera Astivia, O. L., & Ark, T. K. (2015). A methodology for Zumbo's third generation DIF analyses and the ecology of item responding. *Language Assessment Quarterly*, 12(1), 136–151.